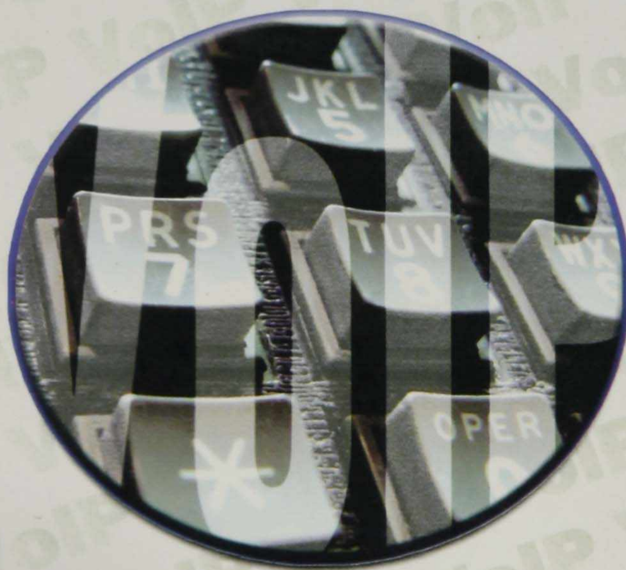
ThS. NGÔ XUÂN THÀNH

CÔNG NGHỆ VoIP

VÀ CÁC GIẢI PHÁP NÂNG CAO CHẤT LƯỢNG DỊCH VỤ





NHÀ XUẤT BẢN BƯU ĐIỆN

CÔNG NGHỆ VoIP
VÀ CÁC GIẢI PHÁP NÂNG CAO CHẤT LƯỢNG DỊCH VỤ

THS. NGÔ XUÂN THÀNH

CÔNG NGHỆ VoIP
VÀ CÁC GIẢI PHÁP
NÂNG CAO CHẤT LƯỢNG DỊCH VỤ

NHÀ XUẤT BẢN BƯU ĐIỆN
Hà Nội, tháng 9 - 2005

LỜI NÓI ĐẦU

Trong những năm gần đây công nghệ truyền thoại đã có nhiều thay đổi vượt bậc. Thoại giờ đây đã có thể truyền trên mạng gói và một trong các công nghệ thoại gói nhiều hứa hẹn đó là thoại IP. Công nghệ truyền thoại qua mạng sử dụng giao thức Internet ra đời với nhiều ưu điểm và mang lại nhiều lợi ích cho cả người sử dụng và nhà khai thác viễn thông. Thoại qua giao thức IP có thể là VoIP hoặc thoại Internet đang ngày càng được sử dụng rộng rãi.

Nhìn chung, các cuộc điện thoại qua giao thức IP có chất lượng thấp hơn so với các cuộc điện thoại truyền thống. Đôi khi, sự khác biệt về chất lượng này là nhỏ và người sử dụng không nhận thấy được. Tuy nhiên, khi thực hiện các cuộc gọi qua Internet thì người dùng sẽ thấy chất lượng thoại giảm một cách rõ rệt. Để thoại qua giao thức IP thực sự đáp ứng được mong đợi của người sử dụng thì vấn đề chất lượng dịch vụ cần giải quyết tốt hơn nữa.

Nhằm cung cấp cho độc giả công nghệ VoIP đến bạn đọc, Nhà xuất bản Bưu điện xuất bản cuốn sách ***"Công nghệ VoIP và các giải pháp nâng cao chất lượng dịch vụ"****. Cuốn sách gồm 6 chương:*

Chương I: Giới thiệu chung, trình bày về xu hướng và ứng dụng của thoại IP cũng như các đặc điểm của tín hiệu thoại.

Chương II: Tổng quan giao thức TCP/IP. Là bộ giao thức được sử dụng cho các mạng IP, mạng sử dụng để truyền thoại IP trên đó.

Chương III: Công nghệ VoIP. Chương này trình bày cấu trúc một hệ thống VoIP và các kỹ thuật mã hóa, truyền tải sử dụng trong thoại IP.

Chương IV: Các giao thức điều khiển, là các bộ giao thức điều khiển sử dụng cho thoại IP.

Chương V: QoS và các yếu tố chính ảnh hưởng đến chất lượng thoại

Chương VI: Giải pháp nâng cao chất lượng thoại. Chương này trình bày một số giải pháp có thể sử dụng để nâng cao chất lượng thoại.

Hi vọng cuốn sách sẽ là tài liệu tham khảo hữu ích cho các cán bộ đang hoạt động trong ngành Bưu điện, các kỹ sư, các kỹ thuật viên đang hoạt động trong lĩnh vực viễn thông, các sinh viên và bạn đọc quan tâm đến vấn đề này.

Ý kiến đóng góp của quý vị và bạn đọc xin gửi về Nhà xuất bản Bưu điện - 18 Nguyễn Du, Hà Nội.

Trân trọng cảm ơn./.

Hà Nội, tháng 8 năm 2005

NHÀ XUẤT BẢN BƯU ĐIỆN

Chương I

GIỚI THIỆU CHUNG

1.1. LỊCH SỬ VÀ XU HƯỚNG PHÁT TRIỂN CỦA THOẠI IP

Công nghệ viễn thông gần đây có nhiều thay đổi lớn, các mạng viễn thông chủ đạo hoạt động trên nền chuyển mạch kênh sẽ có xu hướng chuyển sang hoạt động trên nền chuyển mạch gói. Điều này có nghĩa là phần lớn lưu lượng thoại sẽ được truyền qua mạng chuyển mạch gói. Điện thoại IP ra đời đã gây được sự chú ý mạnh mẽ và có khả năng trở thành nền tảng cho mạng thoại dùng công nghệ chuyển mạch gói. Nhưng liệu điện thoại IP có thể thay thế hoàn toàn điện thoại truyền thống sử dụng công nghệ chuyển mạch kênh không? Hay liệu có còn chỗ cho chuyển mạch kênh khi hiện nay trên thế giới công nghệ chuyển mạch gói đang phát triển mạnh mẽ? Thời gian sẽ trả lời điều đó. Điều mà chúng ta đang nhận thấy là mạng thế hệ mới sử dụng công nghệ chuyển mạch gói đang ngày càng lấn lướt và thu hẹp dần phạm vi của mạng PSTN sử dụng công nghệ chuyển mạch kênh truyền thống và các mạng truyền dữ liệu trước đây.

Khái niệm truyền tín hiệu thoại qua mạng chuyển mạch gói không phải là một khái niệm mới. Nó đã được nghiên cứu khá nhiều vào những năm cuối thập kỷ 1970 và những năm đầu thập kỷ 1980. Cuối thập kỷ 1970 đã có những nghiên cứu và thử nghiệm với thoại gói hóa truyền qua mạng APPANET (tiền thân

của mạng Internet ngày nay) cho thấy điều này có thể thực hiện được. Đầu năm 1995, công ty VocalTec đưa ra sản phẩm phần mềm thoại qua Internet (kết nối điểm - điểm) đầu tiên trên thế giới. Sau đó, nhiều công ty đã đầu tư nghiên cứu và đưa ra sản phẩm thương mại. Tháng 3 năm 1996, VocalTec kết hợp với Dialogic đã đưa ra sản phẩm cổng kết nối PSTN và Internet (gateway) đầu tiên trên thế giới. Hiệp hội các nhà sản xuất thoại qua máy tính (ECTF) đã ra đời nhằm đưa ra các tiêu chuẩn cho thoại qua giao thức Internet. Hiệp hội này bao gồm 36 công ty máy tính và viễn thông hàng đầu thế giới như AT&T, IBM, Sun Microsystem, Digital, Ericsson...

Những đón nhận của các thành phần trong xã hội đối với thoại gói có khác nhau nhưng đều có xu hướng tích cực thúc đẩy nó phát triển.

- Đối với người sử dụng Internet là hộ gia đình thì xem thoại gói như một phương tiện để thực hiện các cuộc gọi từ bất kỳ nơi nào trên thế giới.

- Các nhà quản lý công ty viễn thông xem thoại gói như là phương tiện để tích hợp thoại và dữ liệu trong một mạng duy nhất và nhờ đó mà tiết kiệm giá thành đầu tư và quản lý.

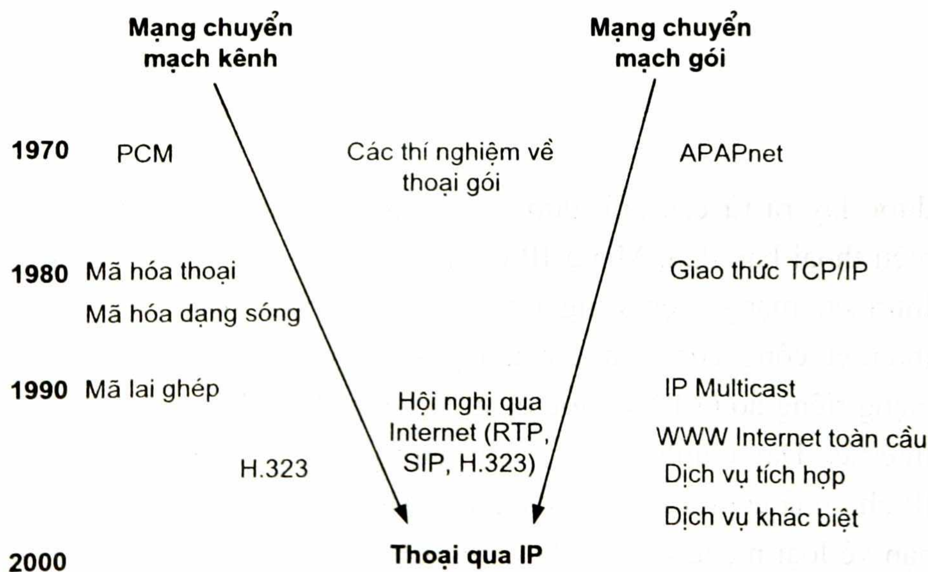
- Các nhà cung cấp dịch vụ Internet (ISP) có thể xem thoại như một dịch vụ và có thể hy vọng tạo ra lợi nhuận từ dịch vụ mới.

- Các công ty cung cấp đường truyền có thể sử dụng hiệu quả hơn đường truyền và có thể chia sẻ tài nguyên đường truyền giữa các mạng thoại và Internet.

- Các nhà sản xuất thiết bị có thể bán thêm nhiều sản phẩm mới

- Các công ty thoại có cơ hội giảm giá thành và tăng tính cạnh tranh.

Hình 1.1 dưới đây là lịch sử phát triển của thoại qua giao thức IP. Các nghiên cứu và thực nghiệm về sự hội tụ của các công nghệ từ những năm của thập kỷ 1980. Tuy nhiên, sự hội tụ mạng chuyển mạch gói và thoại diễn ra chậm, cho mãi đến giữa thập kỷ 1990 các thực nghiệm mới thành công. Và từ đó, thoại truyền qua giao thức IP được phát triển mạnh mẽ.



Hình 1.1: Lịch sử phát triển của VoIP

Với lợi thế giá cước thấp, dịch vụ thoại qua mạng IP thực sự đã làm nhiều nhà kinh doanh trên thế giới quan tâm. Tuy mới được triển khai, song lưu lượng thoại qua mạng Internet đã tăng

đáng kể. Việc đưa VoIP vào mạng có thể coi là một cuộc cách mạng thực sự trong ngành viễn thông.

Thoại IP là công nghệ truyền thoại qua mạng sử dụng giao thức IP, nó có thể gọi là VoIP hoặc thoại Internet. Mạng IP đã phát triển nhanh chóng và ngày càng trở nên phổ biến trong những năm gần đây. Thoại truyền trong mạng IP có nhiều ưu điểm hơn thoại qua mạng chuyển mạch kênh như: dễ dàng tích hợp các loại dịch vụ, thông tin đa phương tiện; sử dụng hiệu quả băng thông.

Động lực chính để phát triển truyền thoại trên nền IP là tiết kiệm băng thông, giảm được giá thành mạng. Nguyên lý của công nghệ này là tín hiệu thoại được số hóa, sau đó phân đoạn (ví dụ phân thành các đoạn với chiều dài là 10 ms) và đóng vào các gói tin sau đó gửi chúng qua mạng IP. Tại đầu thu, dữ liệu được lấy ra từ các gói, được giải mã và biến đổi lại thành tín hiệu thoại ban đầu. Mạng IP có thể là mạng cục bộ (LAN) như Intranet, mạng diện rộng (WAN), mạng pha trộn giữa mạng Internet công cộng và các mạng Extranet logic cá nhân như mạng riêng ảo (VPN) được quản lý bằng IP và lớn nhất là mạng Internet. Đây chính là một trong những môi trường hấp dẫn của IP cho nhà cung cấp dịch vụ mạng. Hoàn toàn không có sự giới hạn về loại mạng mà IP chạy trên đó. IP có thể chạy trên nhiều loại mạng mà không cần những thay đổi lớn về cấu trúc gói tin IP và các cơ chế hoạt động. IP là một giao thức mạng chung, nó có thể phù hợp với rất nhiều loại loại mạng mà không đòi hỏi các phiên bản đặc biệt để thực hiện chức năng chuyển đổi cho phù hợp từ mạng LAN sang mạng WAN và ngược lại.

Chi phí cho dịch vụ điện thoại không phải là rẻ khi so sánh với các chi phí sinh hoạt thông thường trong gia đình và chi phí kinh doanh. Trước đây điện thoại là một dịch vụ rất xa xỉ và chi phí cho nó là khá lớn. Cước phí cho cuộc gọi nội hạt đã cao, nhưng cho cuộc gọi đường dài còn cao hơn và đặc biệt là cuộc gọi quốc tế. Khởi đầu điện thoại chủ yếu phục vụ cho việc kinh doanh, tuy nhiên sau đó nó xâm nhập vào thị trường là các gia đình. Như vậy, vấn đề cước phí được quan tâm hàng đầu. Bốn ưu điểm thường được đề cập đến của VoIP là:

- Giảm cước phí dịch vụ thoại đường dài
- Hỗ trợ nhiều cuộc gọi với băng tần nhỏ hơn
- Khả năng hỗ trợ nhiều dịch vụ mới với chất lượng tốt hơn
- Sử dụng có hiệu quả giao thức IP.

Sau đây chúng ta phân tích từng ưu điểm của VoIP.

1.1.1 Giảm cước phí dịch vụ thoại đường dài

Các đối tượng như doanh nghiệp thường hay gọi các cuộc gọi đường dài nhiều hơn so với hộ gia đình. Vì vậy, giảm cước phí dịch vụ thoại đường dài bao giờ cũng là mục tiêu xuyên suốt. Chi phí gọi đường dài sử dụng VoIP cũng chỉ bằng khoảng 30% chi phí của cuộc gọi qua mạng điện thoại chuyển mạch công cộng (PSTN).

Do đặc thù của kinh doanh, các doanh nghiệp thực hiện rất nhiều các cuộc gọi đường dài. Chúng ta dễ dàng nhận thấy lợi ích lớn nhất của VoIP đối với các doanh nghiệp là giảm cước phí dịch vụ thoại. Trong thực tế, do tính phân tán tự nhiên của các doanh

nghiệp và tính liên kết giữa các doanh nghiệp đã làm cho các cuộc gọi đường dài trở lên rất phổ biến hơn trước đây nhiều.

Các công ty điện thoại nội hạt thường là các công ty độc quyền, có xu thế cố gắng duy trì mức cước để tránh giảm lợi nhuận. Nói cách khác là các công ty độc quyền muốn duy trì cước phí ở mức cao để thu được nhiều lợi nhuận. Trong khi đó, các công ty điện thoại đường dài ở trong môi trường cạnh tranh khốc liệt lại cố gắng kiếm tiền bằng cách giảm thiểu chi phí để tạo lợi thế cạnh tranh với các đối thủ.

Trong mọi trường hợp, VoIP ít nhiều tạo ra sự cạnh tranh giữa các công ty điện thoại nội hạt và các công ty điện thoại đường dài. Và lợi thế này có thúc đẩy các công ty điện thoại cung cấp dịch vụ VoIP không? Vấn đề này hiện vẫn còn đang được tranh cãi sôi nổi. Trong phần này chúng ta chỉ đề cập đến tính cạnh tranh của VoIP với dịch vụ thoại truyền thống trên khía cạnh nó có thể cung cấp các cuộc gọi nội hạt, đường dài đến các cuộc gọi quốc tế với chi phí thấp nhất.

1.1.2 Nhiều cuộc gọi hơn với băng tần nhỏ hơn

Như chúng ta đã biết một kênh thoại số truyền thống yêu cầu tốc độ 64 kbit/s. Vấn đề ở đây là tín hiệu thoại xuất hiện trong suốt thời gian đàm thoại và chiếm toàn bộ băng tần. Đây là bản chất của công nghệ ghép kênh phân chia theo thời gian (TDM) sử dụng trong mạng chuyển mạch kênh. Việc chiếm toàn bộ băng tần trong toàn bộ thời gian xảy ra ngay cả khi một người im lặng khi nghe hoặc thậm chí ngay cả khi cả hai người đặt điện thoại vào trạng thái giữ đường dây. Một trong những

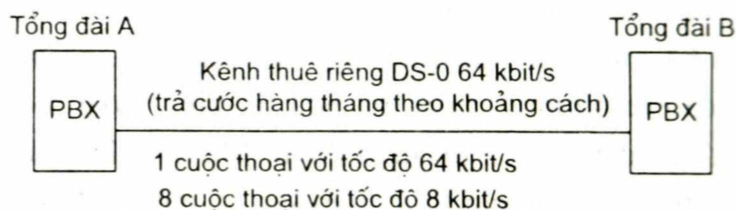
nguyên nhân mà tín hiệu thoại được số hóa ở tốc độ 64 kbit/s là để đảm bảo chất lượng tín hiệu thoại theo định lý Nyquist. Khi công nghệ số hóa tín hiệu thoại được đưa vào sử dụng trong thực tế lần đầu tiên vào đầu những năm 1960 thì số hóa tín hiệu thoại ở tốc độ 64 kbit/s được coi là phù hợp để đảm bảo chất lượng dịch vụ.

Ngày nay, nhờ sự phát triển của công nghệ, điều đó không còn hoàn toàn đúng. Tín hiệu thoại có chất lượng tốt có thể đạt được khi số hóa ở tốc độ nhỏ hơn 64 kbit/s nhiều lần bằng cách sử dụng các phương thức mã hóa hiện đại. Với công nghệ như hiện nay, chất lượng thoại chấp nhận được (chất lượng mà có ít người phàn nàn về dịch vụ thoại) có thể đạt được ở tốc độ cỡ 2 kbit/s, tuy nhiên thường sử dụng tốc độ 8 kbit/s do nhiều nguyên nhân sẽ đề cập sau. Phương pháp mã hóa tín hiệu thoại thấp hơn 64 kbit/s thường được gọi là thoại nén, nhưng không phải tất cả các phương pháp mã hóa tín hiệu thoại tốc độ nhỏ hơn 64 kbit/s đều là nén tín hiệu thoại. Mã hóa tín hiệu thoại ở tốc độ 8 kbit/s thường chỉ thực hiện khi có một ai đó đàm thoại thực sự. Triệt khoảng lặng là cơ chế quyết định của nhiều công nghệ nén.

Các công ty điện thoại thường chia băng tần của mạng thành các kênh thoại tốc độ 64 kbit/s. Các kênh thoại này được gọi là các kênh cơ sở (DS-0) và hiện tại nó vẫn là đơn vị kênh thuê riêng phổ biến nhất. Điều này vẫn đúng khi các kênh thuê riêng DS-0 được sử dụng để truyền dữ liệu như để kết nối các bộ định tuyến (router) ở các vị trí khác nhau. Tất nhiên, người ta có thể thuê nhiều đường E0 để kết nối các tổng đài PBX của họ.

Chúng ta hãy xem xét việc sử dụng kênh thoại E0 để truyền các cuộc thoại giữa các PBX. Một kênh 64 kbit/s chỉ có thể mang được một cuộc thoại. Tuy nhiên nếu thiết bị phía khách hàng (CPE) có khả năng tạo kênh thoại số tốc độ 8 kbit/s, khi đó với cùng độ rộng băng tần có thể kết nối được 8 cuộc thoại trên cùng một kênh. Điều đó tương đương với việc khách hàng đã tăng khả năng kết nối thoại lên 8 lần. Nếu như giá của thiết bị đầu cuối nhỏ hơn cước thuê thêm 7 kênh thuê riêng thì doanh nghiệp đó có động cơ sử dụng công nghệ thoại mới vì đã tiết kiệm được chi phí.

Việc tối ưu hóa sử dụng một kênh DS-0 được thể hiện trên hình 1.2. Nó không có tính thực tế mà chỉ là ví dụ đơn giản để minh chứng cho khái niệm này.

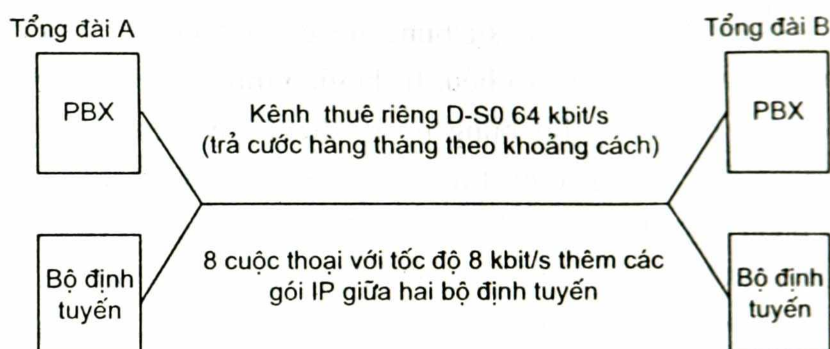


Hình 1.2: Nhiều cuộc gọi hơn với ít băng tần hơn

Việc số hóa tín hiệu thoại ở tốc độ 8 kbit/s như trên vẫn sử dụng toàn bộ băng tần trong toàn bộ khoảng thời gian. Bằng việc sử dụng cơ chế triệt khoảng lặng và số hóa tín hiệu thoại đã tạo ra sự "bùng nổ" dòng bit. Cơ chế này tạo ra các bit thoại số hóa dựa trên sự xuất hiện đột ngột của tín hiệu thoại, thậm chí ngay cả khi một người nói liên tục thì theo cơ chế nén tín hiệu thoại mới cũng chỉ được coi là nhiều lần tín hiệu thoại xuất hiện đột

ngột. Tương tự như vậy, sự bùng nổ các bit cũng là một đặc điểm của các ứng dụng số liệu. IP là giao thức phổ biến nhất để định tuyến các dòng bit bùng nổ từ điểm này đến điểm khác. Chính vì vậy ta gói các bit tín hiệu thoại vào gói IP, các gói tin được phân biệt dựa trên phần mào đầu và nội dung của chúng sau đó gửi chúng đi trên mạng. Đó cũng chính là VoIP sử dụng cơ chế nén và triệt khoảng lặng. Các gói dữ liệu (phi thoại) cũng được gửi đi trong mạng, tuy nhiên sự bùng nổ tín hiệu thoại và dữ liệu ít khi xuất hiện đồng thời. Nếu điều này xảy ra sẽ làm tăng độ trễ các gói tin thoại do kích thước hạn chế của bộ đệm hoặc mất thông tin do phải loại bỏ một số gói. Vì vậy phải sử dụng cơ chế ưu tiên các gói mang thông tin thời gian thực (thoại, video). Tuy nhiên người ta vẫn truyền cả thoại lẫn dữ liệu trên cùng một kênh.

Hình 1.3 minh họa việc sử dụng một kênh DS-0 để truyền không chỉ các cuộc thoại giữa hai PBX của công ty mà còn cả các gói IP giữa các bộ định tuyến. Trong hình này, các tổng đài PBX dùng chung đường kết nối với các bộ định tuyến. Cấu hình ở đây rất đơn giản, tuy nhiên trong thực tế có vấn đề đặt ra là: PBX có các giao diện hoạt động trong mạng chuyển mạch kênh còn các bộ định tuyến IP có các giao diện hoạt động trong mạng chuyển mạch gói. Khi đó, các tổng đài thoại PBX phải có các giao diện hoạt động trong môi trường mạng truyền số liệu. Điều này đưa chúng ta đến ưu điểm thứ 3 là: tạo ra nhiều loại dịch vụ tân tiến và hỗ trợ tốt hơn các dịch vụ bổ sung chủ yếu là cho các dịch vụ số liệu.



Hình 1.3: Kết hợp số liệu và thoại trong VoIP

1.1.3 Hỗ trợ các dịch vụ mới

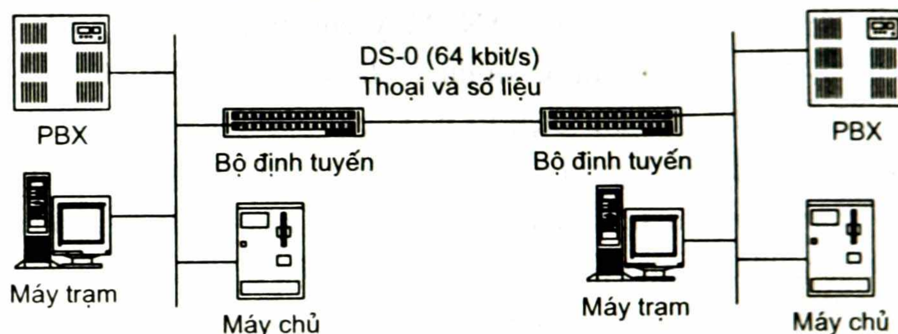
Nói chuyện trên điện thoại chỉ là dịch vụ cơ bản của PSTN. Mạng PSTN cũng hỗ trợ rất nhiều các dịch vụ cộng thêm. Ngày nay nhiều người coi các dịch vụ cộng thêm này là hiển nhiên. Các dịch vụ cộng thêm là các dịch vụ bổ sung vào dịch vụ thoại cơ sở, các dịch vụ này thường phải trả cước phí bổ sung. Một số loại dịch vụ cộng thêm hữu dụng cơ bản là: chuyển hướng cuộc gọi, chờ cuộc gọi, cuộc gọi tay ba, nhận dạng thuê bao chủ gọi (CLIP). Đặc điểm của các dịch vụ cộng thêm này phụ thuộc vào tổng đài và loại báo hiệu mà nó sử dụng.

Tuy nhiên vấn đề ở đây là làm thế nào để chia sẻ kênh giữa PBX và bộ định tuyến IP trong khi PBX có các giao diện và thủ tục tương ứng với mạng chuyển mạch kênh, còn bộ định tuyến IP có các giao diện và thủ tục tương ứng với mạng chuyển mạch gói. Có giải pháp là đưa PBX vào hoạt động trong mạng LAN tương tự như một bộ định tuyến và khi đó các gói tín hiệu thoại được coi giống như các gói số liệu thông thường và được xử lý như các gói số liệu.

Kênh DS-0 sử dụng để truyền cả thoại và số liệu được mô tả trên hình 1.4. Trong cấu hình này, PBX đóng vai trò như là một loại máy chủ trong mạng LAN. Máy chủ trong mạng LAN là một máy tính cá nhân được sử dụng cho chức năng đặc biệt như lưu trữ toàn bộ các thư điện tử hoặc các bản ghi cơ sở dữ liệu cho toàn hệ thống. Người dùng sử dụng các PC đóng vai trò như máy trạm trong mạng LAN, các máy trạm này sử dụng cho các chức năng chung như đọc và gửi các thư điện tử của người dùng, truy nhập đến các bản ghi cơ sở dữ liệu mà người sử dụng quan tâm thông qua một truy vấn cơ sở dữ liệu. Máy trạm và máy chủ được liên kết với nhau thông qua mạng LAN trong đó bộ định tuyến đóng vai trò như phần tử liên kết các vị trí riêng biệt. Bộ định tuyến có thể kết nối mạng LAN giữa các tầng khác nhau trong một toà nhà hoặc kết nối LAN trên toàn thế giới. Với bất kỳ cấu hình kết nối mạng LAN, bất kỳ máy trạm nào đều có thể truy nhập đến mọi máy chủ để lấy thông tin cần thiết và thực hiện một tác vụ nào đó. Cấu hình này được biết đến dưới cái tên khách-chủ và tất cả các ứng dụng chạy trên cấu hình này đều yêu cầu một phiên bản phần mềm chạy trên máy khách sử dụng cùng với phiên bản phần mềm chạy trên máy chủ. Trong mạng có cấu trúc khách-chủ tương ứng với các giao thức sử dụng trong nó yêu cầu tất cả các thiết bị liên kết đều là khách hoặc chủ.

Như vậy, VoIP tạo ra mối quan hệ mật thiết giữa các thiết bị thoại và các thiết bị dữ liệu truyền thống. Các tín hiệu thoại đã nén và triệt khoảng lặng không chỉ được xem là một dạng dữ liệu, mà nó có thể được coi là dữ liệu thực sự trong nhiều trường hợp khác nhau. Thoại VoIP có thể được ghi, lưu trữ, xử lý,

chuyển đổi trên cùng phần cứng và phần mềm mà đã được sử dụng cho các ứng dụng dữ liệu trước kia.



Hình 1.4: PBX đóng vai trò như máy chủ (server) trong mạng LAN

Thiết bị máy chủ VoIP trong mạng LAN có thể hỗ trợ các dịch vụ bổ sung mà các thiết bị PBX truyền thống có thể cung cấp. Ngày nay xu thế tích hợp máy tính và điện thoại (CTI) mà VoIP là một trong những ứng dụng của công nghệ CTI. Với công nghệ CTI thậm chí ngay cả khi không phải là VoIP, một tổng đài PBX có giao diện mạng LAN có thể chuyển số thuê bao chủ gọi không chỉ tới máy điện thoại của một thuê bao A nào đó mà còn có thể tới máy tính của thuê bao này. Một máy trạm có thể tìm kiếm cơ sở dữ liệu trong máy chủ của mạng LAN, sử dụng số thuê bao chủ gọi để hiển thị lên màn hình của nó không chỉ số thuê bao mà còn cả các thông tin liên quan đến thuê bao này ví dụ như tên, ảnh, thông tin về tài khoản, thông tin thẻ tín dụng, tên huý, môn thể thao ưa thích, sở thích... Tất cả các thông tin này xuất hiện trên màn hình trước khi điện thoại rung chuông lần thứ hai.

Tuy nhiên với công nghệ CTI, thiết bị quan trọng là máy tính chứ không phải chỉ có bản thân PBX. Chúng ta hãy xem xét dịch vụ bổ sung khác dựa trên công nghệ CTI. Thay vì nhấc ống nghe điện thoại và quay số, thuê bao A có thể mở một cửa sổ danh bạ trên máy tính của mình và duyệt qua cơ sở dữ liệu các hợp đồng với khách hàng một cách dễ dàng như là duyệt qua danh sách địa chỉ e-mail. Tất cả những gì mà thuê bao A phải làm là chỉ và bấm chuột vào tên người cần liên lạc, số điện thoại của người đó có thể sẽ không xuất hiện như bình thường. Mạng LAN sẽ chuyển số cần gọi đến PBX, PBX sẽ thực hiện cuộc gọi và kết quả là cả hai điện thoại đổ chuông cùng một lúc. Nếu như thuê bao bị gọi bận thì chỉ có chớp trên màn hình và không ảnh hưởng gì đến điện thoại để bàn.

Các dịch vụ chờ cuộc gọi, chuyển hướng cuộc gọi và tất cả các dịch vụ bổ sung có thể cung cấp bởi PBX đều được xử lý một cách tương tự. Một biểu tượng sẽ xử lý tất cả các sự kiện đến và một biểu tượng sẽ thực hiện tất cả các chức năng hướng ra khi ta nhấn chuột vào. CTI còn có thể được sử dụng để nhắc nhở một người gọi điện về nhà trước khi rời văn phòng và thậm chí nó còn tự động thực hiện cuộc gọi cho bạn. Đó là một trong số các tính năng "thông minh" mà công nghệ CTI có thể đem lại.

Nhưng máy chủ VoIP trong mạng LAN hỗ trợ nhiều dịch vụ bổ sung hơn PBX thậm chí ngay cả khi PBX có tích hợp công nghệ CTI. Sở dĩ VoIP đạt được tính năng vượt trội này là do tính phổ biến của các giao thức liên quan đến Internet, Web và các ứng dụng trong môi trường kinh doanh ngày nay. Sự bùng nổ sử

dụng IP sẽ được đề cập đến trong phần tiếp theo. Khi bộ giao thức Internet (TCP/IP) được sử dụng trong mạng LAN thì nó được gọi là mạng Intranet. Hai công ty sử dụng giao thức TCP/IP, một máy trạm ở công ty này có thể đăng ký truy nhập thông tin trong máy chủ của công ty kia và được gọi là Extranet. Chúng ta sẽ đạt được hiệu quả rất cao khi kết hợp Internet và Web để cung cấp các dịch vụ cộng thêm mà các tổng đài PBX thường cung cấp.

1.1.4 Sử dụng IP một cách có hiệu quả nhất

Như đã đề cập đến ở trên, ưu điểm thứ tư cũng là ưu điểm cuối cùng của VoIP sẽ thúc đẩy việc triển khai các thiết bị VoIP.

IP là một giao thức mở trong môi trường mạng Internet và Web. Khoảng giữa những năm 1990, IP không chỉ là giải pháp tốt mà nó còn là giải pháp bắt buộc để mọi người truy nhập Internet và Web nhằm thực hiện các công việc hàng ngày. Mạng Internet và một phần của nó được biết đến là WWW (World Wide Web) không phải là mạng đa giao thức, nó chỉ sử dụng một giao thức đó là IP. Nếu như bạn muốn truy nhập đến một trang Web thì máy tính cá nhân của bạn phải sử dụng giao thức IP, nếu không thì trước hết bạn phải truy nhập đến một cổng kết nối (gateway) chuyên dụng. Việc thử nghiệm các cổng kết nối chuyên dụng là một giải pháp tốt để có thể sử dụng IP ở mọi nơi trong các máy trạm và máy chủ của các công ty. Vì IP là một giao thức mở nên dễ dàng tìm được một nhà cung cấp thiết bị sử dụng các phiên bản của giao thức IP chạy trên hầu hết các nền tảng phần cứng từ các máy MAC đến các máy tính lớn (mainframe).

Khi chuyển sang sử dụng giao thức Internet, các mạng doanh nghiệp trở thành mạng Intranet. Hai điều kiện cần đầu tiên của mạng vẫn tồn tại, đó là nhiều điểm kết nối và nhiều nền tảng phần cứng, còn điều kiện thứ ba là sử dụng nhiều giao thức bị loại bỏ vì bây giờ tất cả đều sử dụng giao thức IP. Khi tất cả đều sử dụng giao thức IP thì chỉ còn hai điều kiện cần để xây dựng mạng Intranet đó là:

- Nhiều vị trí kết nối
- Nhiều nền tảng phần cứng.

Như đã trình bày ở phần trên, một PBX được đưa vào mạng LAN sử dụng VoIP để hỗ trợ không chỉ các dịch vụ cộng thêm của dịch vụ thoại truyền thống cho các máy trạm mà nó còn cung cấp tốt hơn và nhiều hơn các dịch vụ cộng thêm. Sở dĩ người ta sử dụng IP để hỗ trợ các dịch vụ thoại vì IP tồn tại ở mọi nơi.

IP đã chứng minh tính thích ứng của nó với các đối tượng sử dụng dịch vụ phi dữ liệu. Hầu hết các đối tượng sử dụng đều quen thuộc với các nội dung âm thanh và hình ảnh của trang Web cũng như các dữ liệu dạng text của chúng. Các trang Web thường cung cấp hình ảnh về mọi thứ từ hướng dẫn lắp đặt sản phẩm đến bức tranh toàn cảnh. Truyền thông quảng bá trên các trang Web cũng rất phổ biến. Một khách hàng truy nhập đến trang Web của công ty và có thể nói chuyện với đại diện bán hàng hoặc với chuyên gia hỗ trợ kỹ thuật chỉ với một động tác bấm chuột lên phím có ghi dòng chữ, ví dụ như "Hãy nói chuyện với tôi". Nó làm cho người sử dụng cảm thấy thoải mái hơn là

nhấc máy điện thoại, ngay cả khi nó đặt cạnh máy tính của họ và cố gắng tìm kiếm số điện thoại để gọi để nói chuyện với một nhân viên của công ty. Và tất nhiên nhân viên này trả lời điện thoại nhưng không truy nhập đến cùng trang Web mà khách hàng đang truy nhập đến và kết quả là làm tách rời mạng dữ liệu và mạng thoại truyền thống.

Lợi ích cuối cùng của VoIP là sử dụng các giao thức IP có hiệu quả hơn. Nếu như các mạng đều sử dụng giao thức IP thì câu hỏi đặt ra là tại sao chúng ta không sử dụng luôn nó để truyền tín hiệu thoại?

1.2. ĐẶC ĐIỂM CỦA TÍN HIỆU THOẠI

Các nghiên cứu khoa học về ngôn ngữ cho thấy rằng, việc chúng ta có thể nói chuyện với nhau chính là yếu tố làm cho con người về cơ bản khác với các loài vật. Giọng nói được học một cách nhanh chóng sau khi sinh ra, xuất hiện để điều khiển rất nhiều những bộ phận khác nhau trong bộ não. Điều đó cho phép đứa trẻ có thể nói chuyện với cha mẹ chứ không phải nói chuyện với những con vật. Nó mang lại sự ngạc nhiên cho các nhà nghiên cứu, bởi vì những hiểu biết trước đây coi tiếng nói là do tổ chức của bộ não loài người chứ không phải là một cách nào khác. Bộ não con người ngày càng "thông minh" hơn vì họ cần phải thể hiện những suy nghĩ của mình bằng tiếng nói tốt hơn. Chính vì thế mà có thể chỉ loài người là sử dụng tiếng nói như những phương tiện cơ sở cho các trao đổi thông tin xã hội. Và khi nhu cầu liên lạc bằng giọng nói trong xã hội loài người mở rộng ra toàn thế giới, dẫn tới sự phát triển của hệ thống điện

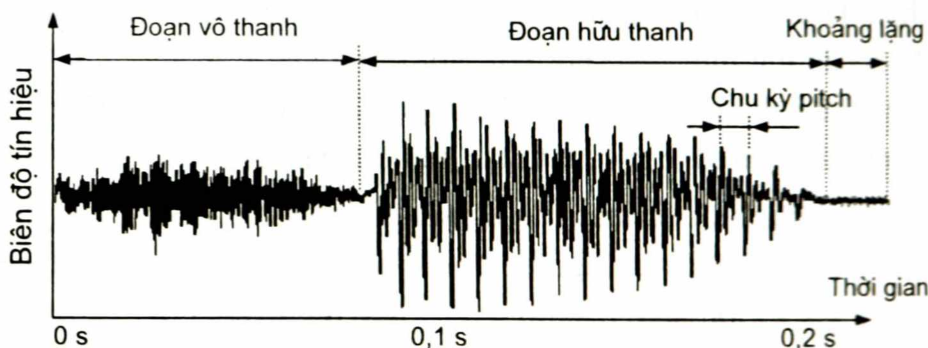
thoại toàn cầu. Hiện nay, với sự phát triển và tiện ích của Internet, người ta đang dần đưa dịch vụ truyền thoại qua mạng Internet vào trong đời sống con người.

Âm thoại hay giọng nói được tạo ra do luồng không khí từ các lá phổi làm rung các dây thanh quản trong cổ họng kết hợp với các chức năng của vòm miệng tạo ra âm thanh và theo thuật ngữ của khoa học ngôn ngữ là các âm vị. Sự rung động xảy ra ở một dải tần (số lần rung trong một giây) do rung động của dây thanh âm qua các thế hệ nó không thay đổi nhiều. Tuy nhiên, dãy các âm thanh lại gây ra sự biến đổi rất nhiều và có nghiên cứu đã chứng tỏ rằng, với một đứa trẻ nếu một âm tiết cụ thể không được học một cách chính xác cho đến khoảng 7 tuổi thì sẽ không có khả năng để phát ra âm đó một cách chính xác. Đây chính là một lý do tại sao khi người ta càng học ngoại ngữ muộn thì họ dường như luôn nói với một trọng âm. Các âm được sắp xếp lại thành một đơn vị âm thanh gọi là từ, các từ được tổ chức thành các cụm từ, các cụm từ tạo nên các câu...

Âm thanh được mang trong sóng âm là một dạng thông tin tương tự. Điều này có nghĩa là biên độ của tín hiệu có thể biến đổi từ giá trị lớn nhất cho tới giá trị nhỏ nhất và có thể là bất cứ giá trị nào trong khoảng đó. Các tín hiệu âm thoại có thể mô tả theo một vài phương thức. Có thể mô tả âm thoại một cách đơn giản là sự pha trộn của các âm có biên độ lớn được gọi là âm hữu thanh (âm tiếng) và các âm có biên độ thấp được gọi là âm vô thanh (âm câm). Các âm hữu thanh bao gồm các nguyên âm như *a, e, i, o* và *u*. Các âm hữu thanh được hình thành trong cổ họng và được quyết định trực tiếp bằng vị trí của dây thanh âm.

Các âm vô thanh là các phụ âm như *b, t, v*... Các âm vô thanh được hình thành từ lưỡi và môi khi không khí đi qua các dây thanh âm. Các âm hữu thanh có biên độ gấp khoảng 20 lần so với các âm vô thanh và mang hầu hết năng lượng của giọng nói. Tuy nhiên, tai dựa vào các âm vô thanh để phân biệt giữa các từ có cùng nguyên âm, như phân biệt giữa các từ: *ta, ma, xa*...

Hình 1.5 biểu diễn biên độ tương quan giữa âm hữu thanh và âm vô thanh trong sóng âm của một đoạn thoại. Hình vẽ minh họa sóng cơ học theo thời gian. Các ngôn ngữ khác nhau sử dụng các âm khác nhau, nhưng tất cả đều được hình thành từ cổ họng và miệng. Dải giới hạn của các âm là một đặc tính quan trọng của thoại tương tự sẽ được áp dụng trong các công nghệ số hóa âm thanh hiện đại.



Hình 1.5: Biểu diễn dạng sóng của một đoạn tín hiệu thoại

Âm hữu thanh, như minh họa trong hình 1.5, có dạng sóng khá đều khi các rung động trong cổ họng ổn định. Mặt khác, các âm vô thanh có dạng sóng ngẫu nhiên không xác định trước bởi vì vị trí của miệng có thể thay đổi trong quá trình phát âm.

Những quan sát này cũng sẽ rất quan trọng trong các công nghệ số hóa thoại.

Các dạng sóng âm có xu hướng lặp lại theo cách được gọi là chu kỳ âm lượng (pitch). Độ dài của chu kỳ âm lượng có thể thay đổi đặc biệt là giữa nam và nữ. Khi người đàn ông nói, chu kỳ âm lượng kéo dài trong khoảng 5 tới 20 ms và khi người đàn bà nói, chu kỳ âm lượng kéo dài trong khoảng 2,5 tới 10 ms.

Đơn vị lý thuyết cơ sở cấu thành nên thoại là âm vị (phoneme). Nó là có chiều dài trung bình khoảng 80 đến 120 ms (đấy là đối với tiếng Anh, tiếng Việt hiện nay chưa có kết quả nghiên cứu nào về vấn đề này). Nếu như thông tin với chiều dài lớn hơn một âm vị thì có thể ảnh hưởng đến khả năng nghe hiểu của từ đó. Do đó một âm hữu thanh đơn có thể có từ 5 tới 50 chu kỳ âm lượng lặp lại. Có điều đặc biệt là không phải tất cả các chu kỳ âm lượng là cần thiết cho nhận dạng giọng nói. Con người có thể hiểu tiếng nói nhanh hơn là họ phát âm. Một vài chương trình quảng cáo trên các chương trình truyền hình, để có thể tăng tốc độ phát tin và tạo các hiệu ứng lạ, thông thường họ bỏ một vài chu kỳ âm lượng lặp lại trong âm tiếng để tăng tốc độ âm thanh mà không làm méo nó. Kết quả âm thanh có thể gây ngạc nhiên cho người nghe nhưng rất ít người có thể nói một cách chính xác sự khác nhau đó là gì.

Trên thực tế, khi ta nói sẽ có khoảng ngắt giữa các câu và khoảng ngắt giữa các từ gọi là khoảng lặng, khoảng lặng đóng vai trò quan trọng trong liên lạc bằng tiếng nói. Tại một thời điểm người nghe hầu như hoàn toàn im lặng trong cuộc đàm

thoại, chỉ có những âm ngắn được phát ra để cho người nói biết rằng họ đang nghe như *ừ*, *vậy hả* hoặc họ vẫn hiểu người nói và làm cho người nói biết họ muốn nghe gì như "*nói tiếp đi*". Những âm phản hồi thường lặp lại khi hai người hội thoại và hoàn toàn không có những âm này khi người được chỉ định nằm trong nhóm người nghe.

Hơn nữa, khoảng lặng thường diễn ra trong trường hợp người nói cần lấy hơi tại cuối một câu hoặc một cụm từ dài. Các khoảng lặng có khoảng thời gian ngắn thường diễn ra trong các cụm từ, giữa các từ và thậm chí là trong chính các từ đó, thường là với các từ có nhiều âm tiết (một âm tiết là sự kết hợp giữa một phụ âm và nguyên âm - được xem như một đơn vị âm thanh).

Trong suốt khoảng thời gian đàm thoại giữa hai người, các âm cần thiết được phát ra từ một người hoặc người kia chiếm khoảng 40% khoảng thời gian. Một nửa thời gian từ người nghe và khoảng 10% là từ các khoảng dừng giữa các câu, các từ và các âm tiết. Trong một cuộc gọi điện thoại, người nghe cũng tập trung vào các nhiễu nền, không chỉ là các nguồn nhiễu như máy thu hình, đài mà còn là các tiếng rè có âm lượng ở mức thấp liên tục. Nhiễu nền này được gọi là âm ngoài. Sự có mặt của âm ngoài trong khoảng lặng cho phép người nghe nhận ra là đường truyền thoại hoạt động vẫn tốt.

Từ những phân tích ở trên cho thấy, giọng nói được đặc tính hóa theo 3 đặc điểm sau:

- Sự pha trộn giữa các âm biên độ cao và các âm biên độ thấp.

- Sự pha trộn của các dạng sóng không đổi có thể xác định trước của âm tiếng và các dạng sóng ngẫu nhiên của âm câm.

- 60% khoảng thời gian đàm thoại hai chiều là khoảng lặng.

Bất cứ công nghệ số hóa âm thanh nào cũng xem xét các đặc tính của âm thanh này và nếu một công nghệ số hóa âm thanh chỉ xem xét tới một đặc tính, còn công nghệ khác xem xét đồng thời cả ba đặc tính, thì ta có thể nói rằng phương pháp số hóa âm thanh sau tốt hơn phương pháp đầu.

Như trên đã trình bày, thoại là một đoạn sóng âm được tạo bởi sự rung động lan truyền trong không khí. Với âm hữu thanh, áp suất không khí từ phổi đẩy ra làm rung dây thanh quản. Tần số của dao động gọi là tần số thay đổi âm lượng (pitch) F_0, nó phụ thuộc vào hình dạng, sức căng của dây thanh quản và áp lực của luồng không khí. Với âm vô thanh, khi kích thích là một luồng khí chuyển động nhanh đi qua một vùng hẹp trong cơ quan phát âm (gồm cả khoang miệng và âm hầu) trong khi thanh môn (glottis) vẫn mở.

Các âm thoại khác nhau được hình thành do sự biến đổi hình dạng của cơ quan phát âm. Do đó, các thuộc tính phổ của tín hiệu thoại cũng thay đổi theo thời gian phụ thuộc vào sự thay đổi hình dạng của cơ quan phát âm. Các thành phần tần số phụ thuộc vào hình dạng, cỡ và vị trí của các khoang hình thành nên cơ quan phát âm. Mỗi một hình dạng của cơ quan phát âm sẽ cho những tần số cộng hưởng khác nhau.

Graham Bell, một giáo viên điếc ở Boston Masachusetts đã phát minh ra điện thoại. Đó là quá trình chuyển thông tin tương

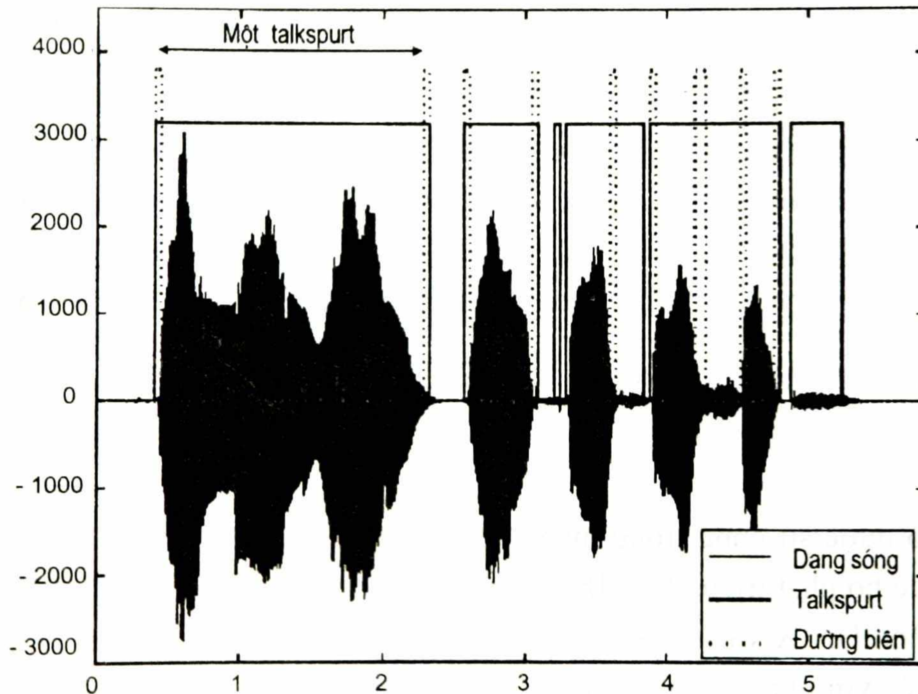
tự (sóng âm thanh) thành các tín hiệu tương tự (sóng điện). Tất cả sự biến đổi biên độ của của âm thanh được thể hiện bởi sóng điện và có thể truyền được đi xa. Bell phát hiện ra rằng để làm cho tiếng nói có thể nghe được trên điện thoại thì dải tần của tiếng nói từ 100 Hz tới 10000 Hz là không thực sự cần thiết. Trên thực tế, chỉ cần một phần nhỏ của dải 10 kHz vì công suất phát ra ở dải trên 4 kHz là rất nhỏ, có đến 80% toàn bộ năng lượng tiếng nói nằm trong dải từ 300 - 3400 Hz.

Và đỉnh (xét về mặt biên độ) được biết đến như formant. Năm formant đầu tiên nằm trong dải tần dưới 3,4 kHz (nó rất quan trọng để nhận dạng thoại, đây cũng là một trong các lý do tại sao người ta lấy dải tần thoại từ 0,3 đến 3,4 kHz). Formant thường xuất hiện như đỉnh trong phổ công suất.

Ta biết rằng, thoại là các đoạn vô thanh, hữu thanh và khoảng lặng nối tiếp nhau liên tiếp như hình 1.6 dưới đây. Các khoảng lặng không mang thông tin gì, cho nên trong thoại IP bộ mã hóa sẽ phát hiện và loại bỏ khoảng lặng. Chỉ những đoạn âm thoại được xử lý, mã hóa và truyền đi.

Trên hình 1.6 cho thấy, vùng màu thẫm bao gồm các đường nét đậm là đường dạng sóng của tín hiệu thoại. Giới hạn giữa đoạn âm thoại và khoảng lặng biểu thị bằng đường chấm chấm. Một đoạn âm thoại (bắt đầu của đoạn âm thoại đến khi kết thúc đoạn âm thoại đó) gọi là một *"talkspurt"*. Mỗi *"talkspurt"* có độ dài ngắn khác nhau, phụ thuộc vào câu và kiểu cách thể hiện cũng như cấu tạo cơ quan phát âm của người nói. Bằng cách phát hiện đoạn tích cực của thoại và bỏ đi những khoảng lặng mà tiết kiệm được đáng kể băng thông. Đồng thời dựa vào các

đặc điểm nêu trên của tiếng nói mà người ta có được các kỹ thuật mã hóa cũng như tái tạo và khôi phục lại tiếng nói trong trường hợp một số gói thoại bị mất để đảm bảo tính trung thực và nguyên vẹn của nó.



Hình 1.6: Dạng sóng của một đoạn thoại

Chương II

TỔNG QUAN GIAO THỨC TCP/IP

2.1. BỘ GIAO THỨC TCP/IP

VoIP hay thoại Internet hoạt động dựa trên giao thức IP. Vì vậy, để hiểu thấu đáo cơ chế của thoại IP trước tiên ta xem xét bộ giao thức TCP/IP (Giao thức điều khiển truyền thông/giao thức Internet – Transmission Control Protocol/Internet Protocol).

2.1.1 Mô hình TCP/IP

TCP/IP là một bộ giao thức được phát triển bởi Cục các dự án nghiên cứu cấp cao (ARPA) của bộ quốc phòng Mỹ. Ban đầu nó được sử dụng trong mạng ARPANET. Khi công nghệ mạng cục bộ phát triển, TCP/IP được tích hợp vào môi trường hệ điều hành UNIX và sử dụng chuẩn Ethernet để kết nối các trạm làm việc với nhau. Đến khi xuất hiện các máy tính cá nhân, TCP/IP lại được chuyển sang môi trường máy tính cá nhân, cho phép hoạt động trên các máy tính cá nhân, hệ điều hành DOS và các trạm làm việc chạy hệ điều hành UNIX có thể liên tác trên cùng một mạng. Hiện nay, TCP/IP được sử dụng rất phổ biến trong mạng máy tính, mà điển hình là mạng Internet.

TCP/IP được phát triển trước mô hình OSI. Do đó, các tầng trong TCP/IP không tương ứng hoàn toàn với các tầng trong mô hình OSI. Chồng giao thức TCP/IP được chia thành bốn tầng:

giao diện mạng (network interface), liên mạng (internet), giao vận (transport) và ứng dụng (application).

MÔ HÌNH OSI	MÔ HÌNH TCP/IP
ỨNG DỤNG	ỨNG DỤNG
TRÌNH DIỄN	
PHIÊN	
GIAO VẬN	GIAO VẬN
MẠNG	LIÊN MẠNG
LIÊN KẾT DỮ LIỆU	GIAO DIỆN MẠNG
VẬT LÝ	

Hình 2.1: Mô hình OSI và TCP/IP

Tầng ứng dụng trong mô hình TCP/IP tương ứng với ba tầng trong mô hình OSI là tầng ứng dụng, tầng trình diễn và tầng phiên. Tầng này còn được gọi là tầng xử lý (process). Tầng giao vận tương ứng với tầng giao vận trong mô hình OSI. Tầng này còn được gọi là tầng trạm-tới-trạm (host-to-host). Tầng liên mạng tương ứng với tầng mạng trong mô hình OSI. Tầng giao diện mạng tương ứng với tầng liên kết dữ liệu và vật lý trong mô hình OSI.

2.1.2 Tầng ứng dụng

Tầng ứng dụng cung cấp các dịch vụ dưới dạng các giao thức cho ứng dụng của người dùng. Một số giao thức tiêu biểu tại tầng này gồm:

- *FTP* (File Transfer Protocol): Đây là một dịch vụ hướng kết nối và tin cậy, sử dụng TCP để cung cấp truyền tệp giữa các hệ thống hỗ trợ FTP.
- *Telnet* (TErminaL NETwork): Cho phép các phiên đăng nhập từ xa giữa các máy tính. Do Telnet hỗ trợ chế độ văn bản nên giao diện người dùng thường ở dạng dấu nhắc lệnh tương tác. Chúng ta có thể đánh lệnh và các thông báo trả lời sẽ được hiển thị.
- *HTTP* (Hyper Text Transfer Protocol): Trao đổi các tài liệu siêu văn bản để hỗ trợ Web.
- *SMTP* (Simple Mail Transfer Protocol): Truyền thư điện tử giữa các máy tính. Đây là dạng đặc biệt của truyền tệp được sử dụng để gửi các thông báo tới một máy chủ thư hoặc giữa các máy chủ thư với nhau.
- *POP3* (Post Office Protocol): Cho phép lấy thư điện tử từ hộp thư trên máy chủ.
- *DNS* (Domain Name System): Chuyển đổi tên miền thành địa chỉ IP. Giao thức này thường được các ứng dụng sử dụng khi người dùng ứng dụng này dùng tên chứ không dùng địa chỉ IP.
- *DHCP* (Dynamic Host Configuration Protocol): Cung cấp các thông tin cấu hình động cho các trạm, chẳng hạn như gán địa chỉ IP.
- *SNMP* (Simple Network Management Protocol): Được sử dụng để quản trị từ xa các thiết bị mạng chạy TCP/IP. SNMP thường được thực thi trên các trạm của người quản lý, cho phép người quản lý tập trung nhiều chức năng giám sát và điều khiển trong mạng.

2.1.3 Tầng giao vận

Tầng giao vận chịu trách nhiệm chuyển phát toàn bộ thông báo từ tiến trình-tới-tiến trình. Tại tầng này có hai giao thức là TCP (Transmission Control Protocol) và UDP (User Datagram Protocol). Mỗi giao thức cung cấp một loại dịch vụ giao vận khác nhau: hướng kết nối và phi kết nối.

Một máy tính có thể có nhiều người dùng và mỗi người dùng có thể có nhiều kết nối TCP hoặc UDP cùng lúc. Làm thế nào có thể giữ được các kết nối này? Điều này được thực hiện bằng cách sử dụng các giao thức tầng giao vận.

Các giao thức tầng giao vận cung cấp các dịch vụ cho phép nhiều tiến trình truyền thông tồn tại trên một máy. Mỗi tiến trình này được gọi là một cổng (port). Trong một mạng TCP/IP, địa chỉ IP được sử dụng để nhận biết một trạm. Tuy nhiên trên một trạm lại có thể có nhiều kết nối TCP/IP cùng lúc, do đó nếu chỉ sử dụng địa chỉ IP thì chưa thể nhận diện được các kết nối TCP/IP. *Socket* (sự kết hợp của địa chỉ IP và cổng) được sử dụng để nhận biết các kết nối TCP/IP. Chúng ta có thể có nhiều *socket* trên một trạm. Mỗi *socket* có thể được một giao thức lớp ứng dụng sử dụng. Do số lượng cổng trên một máy nhiều nên trên một máy có thể có nhiều người dùng truyền thông cùng lúc.

2.1.3.1 Giao thức TCP

Một giao thức tầng giao vận thường có nhiều chức năng. Một trong số đó là tạo một truyền thông tiến trình-tới-tiến trình (chương trình-tới-chương trình). Như đã trình bày, để thực hiện điều này, TCP sử dụng "cổng" (port). Một chức năng khác của

tầng giao vận là tạo một cơ chế điều khiển luồng và điều khiển lỗi ở mức giao vận. TCP sử dụng cơ chế cửa sổ trượt để thực hiện điều khiển luồng. Nó sử dụng gói xác nhận, thời gian chờ và truyền lại để thực hiện điều khiển lỗi.

TCP là một giao thức hướng kết nối. Nó có nhiệm vụ thiết lập một kết nối với phía nhận, chia luồng dữ liệu thành các đơn vị có thể vận chuyển, đánh số chúng và sau đó lần lượt gửi chúng đi.

** Truyền thông tiến trình-tới-tiến trình*

IP có trách nhiệm truyền thông ở mức máy tính (truyền thông trạm-tới-trạm). Là một giao thức tầng mạng, IP chỉ có thể chuyển phát các thông báo tới máy đích. Tuy nhiên, đây chưa phải là một chuyển phát hoàn chỉnh. Thông báo cần được xử lý bởi đúng chương trình ứng dụng. Trách nhiệm chuyển thông báo tới chương trình ứng dụng thích hợp là chức năng của TCP.

** Địa chỉ cổng*

Mặc dù có một số cách để thực hiện truyền thông tiến trình-tới-tiến trình, nhưng cách thông dụng nhất là thực hiện thông qua mô hình khách-chủ (client-server). Một tiến trình trên máy cục bộ, được gọi là khách, cần một dịch vụ từ một ứng dụng trên trạm ở xa, được gọi là chủ.

Cả hai tiến trình (khách, chủ) có cùng một tên. Ví dụ, để lấy thời gian và ngày tháng từ một máy chủ ở xa, chúng ta cần một tiến trình khách Daytime chạy trên máy cục bộ và một tiến trình chủ Daytime chạy trên máy chủ ở xa.

Bây giờ, chúng ta đã biết rằng địa chỉ IP và số cổng đóng vai trò khác nhau trong việc chọn đích cuối cùng của dữ liệu. Địa chỉ IP đích xác định trạm trong số nhiều trạm khác nhau. Sau khi trạm đã được chọn, số cổng xác định một tiến trình trên trạm cụ thể đó.

Các số cổng được chia thành ba vùng: thông dụng, đăng ký và động.

- *Cổng thông dụng*. Các cổng nằm trong khoảng từ 0 đến 1023 là các cổng thông dụng. Những cổng này được gán và giám sát bởi IANA.

- *Cổng đăng ký*. Các cổng nằm trong khoảng từ 1024 đến 49151 không do IANA gán và điều khiển. Chúng chỉ có thể được đăng ký với IANA để tránh trùng lặp.

- *Cổng động*. Các cổng nằm trong khoảng từ 49152 đến 65535 có thể được sử dụng bởi mọi tiến trình. Chúng còn được gọi là các cổng ngẫu nhiên.

** Địa chỉ socket*

Để thiết lập kết nối, TCP cần hai số hiệu nhận dạng: địa chỉ IP và số cổng. Sự kết hợp địa chỉ IP và số cổng được gọi là địa chỉ socket. Để sử dụng dịch vụ TCP, chúng ta cần một cặp địa chỉ socket: địa chỉ socket khách và địa chỉ socket chủ. Địa chỉ socket khách để định danh duy nhất ứng dụng khách. Địa chỉ socket chủ để định danh duy nhất ứng dụng chủ. Bốn thông tin này là một phần của tiêu đề IP và tiêu đề TCP. Tiêu đề IP chứa địa chỉ IP; tiêu đề TCP chứa địa chỉ cổng.

** Phân đoạn TCP*

Đơn vị dữ liệu truyền giữa hai thiết bị sử dụng TCP được gọi là phân đoạn (segment). Định dạng của phân đoạn TCP được cho thấy ở hình 2.2.

Bit 0-3	Bit 4-7	Bit 8-11	Bit 12-15	Bit 16-19	Bit 20-23	Bit 24-27	Bit 28-31
Cổng nguồn				Cổng đích			
Số trình tự							
Số xác nhận							
Chiều dài tiêu đề	Dự phòng	Các cờ		Kích thước cửa sổ			
Tổng kiểm tra				Con trỏ khẩn			
Tùy chọn và đệm							

Hình 2.2: Phần tiêu đề TCP

Phân đoạn gồm một phần tiêu đề có chiều dài từ 20 byte đến 60 byte, theo sau là dữ liệu từ chương trình ứng dụng. Tiêu đề có chiều dài 20 byte nếu nó không chứa tùy chọn và có chiều dài tối đa 60 byte nếu nó chứa một số tùy chọn.

Các trường của phần tiêu đề:

- *Cổng nguồn (Source Port)*: Trường 16 bit này xác định số cổng của chương trình ứng dụng gửi.

- *Cổng đích (Destination Port)*: Trường 16 bit này xác định số cổng của chương trình ứng dụng nhận.

- *Số trình tự (Sequence number)*: Trường 32 bit này xác định số được gán cho byte dữ liệu đầu tiên chứa trong phân

đoạn. Như chúng ta đã nói ở phần trước, TCP là một giao thức giao vận luồng. Để đảm bảo tính kết nối, mỗi byte được gửi đi phải được đánh số. Số trình tự nói cho đích biết số hiệu byte đầu tiên trong phân đoạn. Trong giai đoạn thiết lập kết nối, mỗi phía sử dụng một bộ tạo số ngẫu nhiên để tạo một số trình tự khởi đầu (ISN), số này thường khác nhau trong mỗi hướng. Ví dụ, nếu số ISN là 2367 và gói thứ nhất mang 1000 byte dữ liệu, số trình tự là 2369 (2367 và 2368 được sử dụng để thiết lập kết nối); phân đoạn thứ hai mang 500 byte dữ liệu sẽ có số trình tự 3369 (2369+1000) .v.v...

- *Số xác nhận (Acknowledgment Number)*: Số 32 bit này xác định số hiệu byte mà trạm gửi phân đoạn đang chờ để nhận. Nếu nó đã nhận thành công byte số x, thì số xác nhận của nó là x + 1.

- *Chiều dài tiêu đề (HL – Header Length)*: Trường 4 bit này cho biết chiều dài tính theo từ (4 byte) của phần tiêu đề TCP.

- *Dự phòng*: Trường 6 bit này được dùng cho tương lai.

- *Các cờ*: Trường này xác định 6 bit điều khiển khác nhau. Một hoặc nhiều bit này có thể được đặt tại cùng thời điểm. Sáu bit cờ lần lượt là:

  + *URG*: Khi cờ này được đặt, trường con trỏ khẩn sẽ có hiệu lực. Khi cờ này không được đặt, trường con trỏ khẩn được bỏ qua.

  + *ACK*: Cờ này được đặt để cho biết giá trị của trường xác nhận là hợp lệ.

  + *PSH*: Khi cờ này được đặt, toàn bộ dữ liệu trong bộ đệm (kể cả dữ liệu được lưu từ trước) sẽ được chuyển ngay

lên cho chương trình ứng dụng. Còn nếu cờ này không được đặt, TCP chờ đến khi thích hợp mới chuyển dữ liệu đi, nhằm tăng hiệu quả của hệ thống.

+ *RST*: Cờ này được đặt để báo cho bên nhận biết bên gửi đang từ bỏ kết nối.

+ *SYN*: Cờ này được đặt để đồng bộ hóa các hiệu trình tự.

+ *FIN*: Cờ này được đặt để báo bên gửi không còn dữ liệu.

- *Kích thước cửa sổ (Window Size)*: Trường 16 bit này xác định kích thước của cửa sổ (tính theo byte) mà phía kia phải duy trì. Kích thước tối đa của cửa sổ là 65535. Chúng ta sẽ tìm hiểu chi tiết hơn về trường này ở phần cửa sổ trượt.

- *Tổng kiểm tra (Checksum)*: Trường 16 bit này chứa mã kiểm tra lỗi (theo phương pháp CRC) cho toàn bộ phân đoạn (cả tiêu đề và dữ liệu).

- *Con trỏ khẩn (Urgent Pointer)*: Trường 16 bit này chỉ hợp lệ khi cờ URG được đặt. Nó xác định số phải cộng với số trình tự để lấy được số hiệu của byte khẩn cuối cùng trong phần dữ liệu.

- *Tùy chọn (Option):* Trường có chiều dài tối đa 40 byte này chứa các thông tin tùy chọn.

2.1.3.2 Giao thức UDP

UDP (User Datagram Protocol) là một giao thức truyền thông phi kết nối và không tin cậy, được dùng thay thế cho TCP ở trên IP theo yêu cầu của ứng dụng. UDP có trách nhiệm truyền các thông báo từ tiến trình-tới-tiến trình, nhưng không cung cấp các cơ chế giám sát và quản lý.

UDP cũng cung cấp cơ chế gán và quản lý các số cổng để định danh duy nhất cho các ứng dụng chạy trên một trạm của mạng. Do ít chức năng phức tạp nên UDP có xu thế hoạt động nhanh hơn so với TCP. Nó thường được dùng cho các ứng dụng không đòi hỏi độ tin cậy cao trong giao vận như thoại, video. Định dạng của gói tin UDP được mô tả trong hình 2.3, với các vùng tham số đơn giản hơn nhiều so với phân đoạn TCP.

Bit 0-3	Bit 4-7	Bit 8-11	Bit 12-15	Bit 16-19	Bit 20-23	Bit 24-27	Bit 28-31
Cổng nguồn				Cổng đích			
Độ dài tổng				Tổng kiểm tra			
Dữ liệu							

Hình 2.3: Định dạng của gói tin UDP

Các trường trong tiêu đề gói tin UDP gồm:

- *Cổng nguồn (Source Port)*: Trường 16 bit này xác định số cổng của chương trình ứng dụng gửi.

- *Cổng đích (Destination Port)*: Trường 16 bit này xác định số cổng của chương trình ứng dụng nhận.

- *Độ dài tổng (Total Length)*: Trường 16 bit này xác định độ dài tổng (cả tiêu đề và dữ liệu) của gói tin UDP.

- *Tổng kiểm tra (Checksum)*: Trường 16 bit này chứa mã kiểm tra lỗi (theo phương pháp CRC) cho toàn bộ phân đoạn (cả tiêu đề và dữ liệu).

Một số giao thức tiêu biểu tầng ứng dụng sử dụng UDP gồm:

- Giao thức truyền tệp thông thường (TFTP - Trivial File Transfer Protocol)

- Giao thức quản lý mạng đơn giản (SNMP - Simple Network Management Protocol)

- Hệ thống tệp mạng (NFS - Network File System)

- Hệ thống tên miền (DNS - Domain Name System)

- Giao thức truyền thời gian thực (RTP - Realtime Transfer Protocol)

2.1.4 Tầng liên mạng

Tầng liên mạng trong chồng giao thức TCP/IP tương ứng với tầng mạng trong mô hình OSI. Chức năng chính của tầng liên mạng là đánh địa chỉ lô-gíc và định tuyến gói tin tới đích. Giao thức đáng chú ý nhất ở tầng liên mạng chính là giao thức Internet (IP - Internet Protocol). Ngoài ra còn có một số giao thức khác như ICMP, ARP và RARP. Chúng ta sẽ lần lượt tìm hiểu các giao thức này.

2.1.4.1 Giao thức IP

IP là một giao thức phi kết nối và không tin cậy. Nó cung cấp dịch vụ chuyển gói "cố gắng tối đa". "Cố gắng tối đa" ở đây có nghĩa IP không cung cấp chức năng theo dõi và kiểm tra lỗi. Nó chỉ cố gắng chuyển gói tới đích chứ không có sự đảm bảo. Nếu độ tin cậy là yếu tố quan trọng, IP phải hoạt động với một giao thức tầng trên tin cậy, chẳng hạn TCP.

IP cũng là một dịch vụ phi kết nối, được thiết kế cho một mạng chuyển mạch gói. Phi kết nối có nghĩa mỗi gói dữ liệu được xử lý độc lập, mỗi gói có thể đi tới đích trên một đường đi khác nhau, chúng có thể đến sai thứ tự. Một số gói dữ liệu có thể

bị mất, bị hỏng trong khi truyền. IP dựa vào một giao thức tầng cao hơn để xử lý những vấn đề này.

Các gói dữ liệu tại tầng IP được gọi là gói dữ liệu. Hình 2.4 cho thấy định dạng gói dữ liệu. Một gói dữ liệu có chiều dài biến thiên, gồm hai phần: tiêu đề và dữ liệu. Phần tiêu đề có chiều dài từ 20 đến 60 byte, chứa các thông tin cần thiết cho định tuyến và chuyển phát dữ liệu.

Bit 0-3	Bit 4-7	Bit 8-10	Bit 11-15	Bit 16-19	Bit 20-23	Bit 24-27	Bit 28-i31
Phiên bản	Chiều dài tiêu đề	Độ ưu tiên	Loại dịch vụ	Độ dài tổng			
Số hiệu gói dữ liệu				Phân mảnh			
Thời gian sống		Giao thức		Tổng kiểm tra			
Địa chỉ nguồn							
Địa chỉ đích							
Tùy chọn							

Hình 2.4: Tiêu đề gói dữ liệu IP

- *Phiên bản (Version)*: Trường 4 bit này cho biết phiên bản IP tạo phần tiêu đề này. Phiên bản hiện tại là 4. Tuy nhiên phiên bản IPv6 sẽ thay thế IPv4 trong tương lai.

- *Chiều dài tiêu đề (HL - Header Length)*: Trường 4 bit này cho biết chiều dài của phần tiêu đề gói dữ liệu IP, tính theo đơn vị từ (32 bit). Trường này là cần thiết vì chiều dài của phần tiêu đề thay đổi (từ 20 đến 60 byte). Khi không có phần tùy chọn (option), chiều dài phần tiêu đề là 20 byte và giá trị của trường này là 5 ($5 \times 4 = 20$). Khi phần tùy chọn có kích thước tối đa thì giá trị của trường là 15 ($15 \times 4 = 60$).

- *Độ ưu tiên (Precedence)*: Trường này có chiều dài 3 bit, giá trị nằm trong khoảng từ 0 (000) đến 7 (111). Nó chỉ rõ độ ưu tiên của gói dữ liệu trong trường hợp mạng có tắc nghẽn. Nếu một bộ định tuyến bị tắc nghẽn và cần bỏ một số gói dữ liệu, nó sẽ bỏ các gói dữ liệu có độ ưu tiên thấp nhất.

- *Loại dịch vụ (TOS)*: Trường 5 bit này đặc tả các tham số về dịch vụ, có dạng cụ thể như sau:

Loại dịch vụ				
D	T	R	C	x

Trong 5 bit này, có một bit dự phòng (x). Cho dù các bit đều có thể lấy giá trị 0 hoặc 1, nhưng trong mỗi gói dữ liệu chỉ có một bit được đặt là 1.

+ 00000: Bình thường

+ 00010: Giá nhỏ nhất

+ 00100: Độ tin cậy cao nhất

+ 01000: Thông lượng cao nhất

+ 10000: Độ trễ nhỏ nhất

- *Độ dài tổng (Total Length IP)*: Trường 16 bit này cho biết chiều dài tính theo byte của cả gói dữ liệu.

- *Số hiệu gói dữ liệu (Datagram IDentification)*: Trường 16 bit này cùng với các trường khác (như địa chỉ nguồn và địa chỉ đích) dùng để định danh duy nhất cho một gói dữ liệu trong khoảng thời gian nó vẫn tồn tại trên liên mạng. Giá trị này được tăng lên 1 đơn vị mỗi khi có gói dữ liệu được trạm gửi đi. Do vậy giá trị này sẽ quay lại 0 mỗi khi trạm đã gửi 65535 gói dữ liệu.

- *Phân mảnh (Fragmentation)*: Trường 16 bit này được sử dụng khi gói dữ liệu được phân mảnh (sẽ được trình bày ở phần sau).

- *Thời gian sống (Time to Live)*: Trường 8 bit này qui định thời gian tồn tại (tính bằng giây) của gói dữ liệu trong liên mạng để tránh tình trạng gói dữ liệu bị chuyển vòng quanh trên liên mạng. Thời gian này do trạm gửi đặt và bị giảm đi 1 mỗi khi gói dữ liệu qua một bộ định tuyến trên liên mạng.

- *Giao thức (Protocol)*: Trường 8 bit này cho biết giao thức tầng trên sử dụng dịch vụ của tầng IP. Gói dữ liệu IP có thể đóng gói dữ liệu từ nhiều giao thức tầng trên, chẳng hạn TCP, UDP và ICMP. Trường này chỉ rõ giao thức đích cuối cùng mà gói dữ liệu IP phải chuyển.

- *Tổng kiểm tra (Checksum)*: Trường 16 bit này chứa mã kiểm tra lỗi theo phương pháp CRC (chỉ kiểm tra phần tiêu đề).

- *Địa chỉ nguồn (Source Address)*: Trường 32 bit này chứa địa chỉ IP của trạm nguồn.

- *Địa chỉ đích (Destination Address)*: Trường 32 bit này chứa địa chỉ IP của trạm đích.

2.1.4.2 Giao thức ICMP

Để cho phép các bộ định tuyến trong Internet thông báo lỗi hoặc cung cấp các thông tin về những tình huống không mong đợi, các nhà thiết kế hệ thống đã thêm vào một cơ chế thông báo có mục đích đặc biệt cho các giao thức TCP/IP. Cơ chế này thể hiện qua "giao thức thông báo điều khiển liên mạng" (ICMP - Internet Control Message Protocol), nó được xem như là một

phần bắt buộc của giao thức IP. Giống như các loại bản tin khác, ICMP được chuyển qua Internet trong phần dữ liệu của gói tin IP. Tuy nhiên, đích đến cuối cùng của bản tin này là phần mềm IP và được xử lý để xác định nguyên nhân lỗi.

Sở dĩ cần đến giao thức ICMP vì IP là giao thức chuyển gói phi kết nối và không tin cậy, nó được thiết kế nhằm mục đích sử dụng có hiệu quả tài nguyên mạng. IP cung cấp dịch vụ chuyển gói nỗ lực tối đa, tuy nhiên nó có hai thiếu hụt: thiếu điều khiển lỗi và thiếu các cơ chế hỗ trợ.

Giao thức IP không có cơ chế thông báo lỗi và sửa lỗi. Điều gì xảy ra nếu bộ định tuyến phải bỏ một gói dữ liệu do không tìm thấy bước nhảy tiếp theo cho gói dữ liệu đó, hoặc khi giá trị trường *thời gian sống* bằng không? Điều gì xảy ra nếu trạm đích phải bỏ tất cả các mảnh của gói dữ liệu do không nhận được đủ các các mảnh trong một khoảng thời gian định trước? Đấy là những ví dụ về tình trạng xảy ra lỗi, nhưng IP không có những cơ chế được xây dựng sẵn để thông báo lỗi cho trạm nguồn.

IP cũng thiếu cơ chế truy vấn. Một trạm đôi khi cần xác định xem bộ định tuyến hoặc một trạm khác có hoạt động không và một người quản lý mạng đôi khi cần thông tin từ một trạm hoặc bộ định tuyến khác.

Giao thức thông báo điều khiển liên mạng (ICMP – Internet Control Message Protocol) được thiết kế để bù đắp hai thiếu hụt trên. Nó được đi kèm với giao thức IP. Một bản tin ICMP di chuyển trong phần dữ liệu của gói tin IP và có ba trường dữ liệu có độ dài cố định nằm trong phần đầu của bản tin: Trường kiểu bản tin ICMP, trường mã và trường tổng kiểm tra.

Về mặt kỹ thuật, giao thức ICMP là một cơ chế thông báo lỗi. Nó cung cấp cho các bộ định tuyến một phương pháp để khi gặp lỗi thì sẽ thông báo lỗi cho trạm nguồn đầu tiên, vì chỉ có thông tin về địa chỉ nguồn phát là ổn định còn các bộ định tuyến trung gian sẽ thay đổi vì các bảng định tuyến sẽ luôn được cập nhật theo trạng thái của mạng. Như vậy, giao thức ICMP cung cấp thông tin điều khiển trong quá truyền thông giữa các bộ định tuyến và máy tính; nó là một phần tổng hợp bắt buộc của IP.

** Các loại thông báo ICMP*

Các thông báo ICMP được chia làm hai loại: *thông báo lỗi (error-reporting)* và *truy vấn (query).*

Thông báo lỗi thông báo những sự cố mà bộ định tuyến hoặc trạm đích có thể gặp phải khi xử lý gói dữ liệu IP. *Truy vấn* giúp một trạm hoặc một người quản lý mạng lấy các thông tin cụ thể về một bộ định tuyến hoặc một trạm khác.

2.1.4.3 Giao thức ARP và RARP

Một liên mạng là sự kết hợp của nhiều mạng vật lý được kết nối với nhau thông qua các thiết bị liên kết mạng, chẳng hạn bộ định tuyến hoặc cổng kết nối.

Trạm và bộ định tuyến được nhận dạng tại tầng mạng bằng địa chỉ lô-gíc (địa chỉ IP). Địa chỉ lô-gíc là địa chỉ chung, nó phải là duy nhất trong toàn bộ liên mạng. Tuy nhiên, trên đường tới đích, gói phải đi qua các mạng vật lý khác nhau. Ở mức vật lý, trạm và bộ định tuyến được nhận dạng bởi địa chỉ vật lý (địa chỉ MAC).

Địa chỉ vật lý và địa chỉ IP là hai số hiệu nhận dạng khác nhau. Chúng ta cần cả hai vì một mạng vật lý, chẳng hạn mạng

Ethernet có thể được sử dụng đồng thời bởi hai giao thức tầng mạng khác nhau, chẳng hạn IP và IPX. Ngược lại, một gói thuộc một giao thức tầng mạng, chẳng hạn IP, có thể đi qua nhiều mạng vật lý.

Nghĩa là, để chuyển phát gói tới một trạm hoặc một bộ định tuyến, cần có hai mức đánh địa chỉ: lô-gíc và vật lý. Do vậy, chúng ta cần ánh xạ giữa hai địa chỉ này. Giao thức phân giải địa chỉ (ARP – Address Resolution Protocol) chuyển đổi địa chỉ lô-gic thành địa chỉ vật lý. Và giao thức phân giải địa chỉ ngược (RARP – Reverse Address Resolution Protocol) chuyển đổi địa chỉ vật lý thành địa chỉ lô-gic.

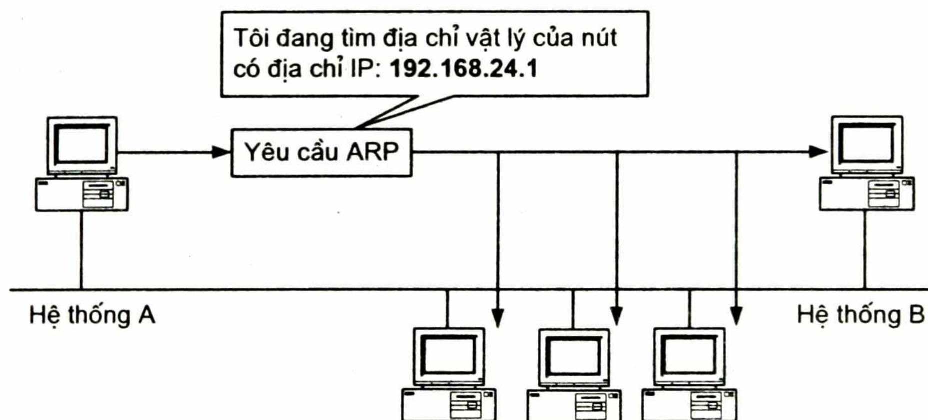
Cùng với giao thức phân giải địa chỉ là giao thức phân giải địa chỉ ngược. RARP yêu cầu phải có một máy chủ RARP, máy chủ trả lời bằng cách điền vào địa chỉ đích thay đổi bản tin trở thành bản tin trả lời với lời đáp trực tiếp, có thể có rất nhiều máy chủ trả lời, máy sẽ nhận lời đáp đầu tiên. Không thể quảng bá quá nhiều thông tin RARP trên mạng vì như vậy có thể gây ra tắc nghẽn hệ thống. Thông tin định tuyến được yêu cầu từ mạng để dẫn các gói tin đến đích cuối cùng của nó nằm trong tiêu đề.

** Giao thức ARP*

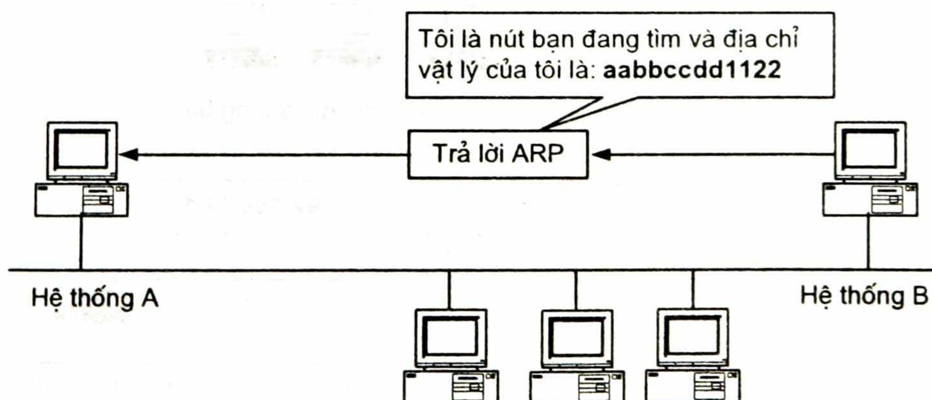
Khi một trạm hoặc bộ định tuyến cần tìm địa chỉ vật lý của một trạm hoặc một bộ định tuyến khác trên mạng, nó gửi gói yêu cầu ARP. Gói này chứa địa chỉ vật lý và địa chỉ lôgic của nguồn và địa chỉ IP của đích. Do nguồn không biết địa chỉ vật lý của đích nên yêu cầu này được gửi quảng bá.

Mọi trạm và bộ định tuyến trên mạng đều nhận và xử lý yêu cầu ARP này, nhưng chỉ có trạm đích nhận ra địa chỉ IP của nó

và gửi trả lời ARP lại cho nguồn. Gói trả lời chứa địa chỉ lô-gíc và địa chỉ vật lý của đích. Gói trả lời này được gửi thẳng (gửi đơn hướng - unicast) tới trạm yêu cầu (nguồn) sử dụng địa chỉ vật lý có trong gói yêu cầu ARP. Hình 2.5 minh họa hoạt động của giao thức ARP



a. Yêu cầu ARP được quảng bá



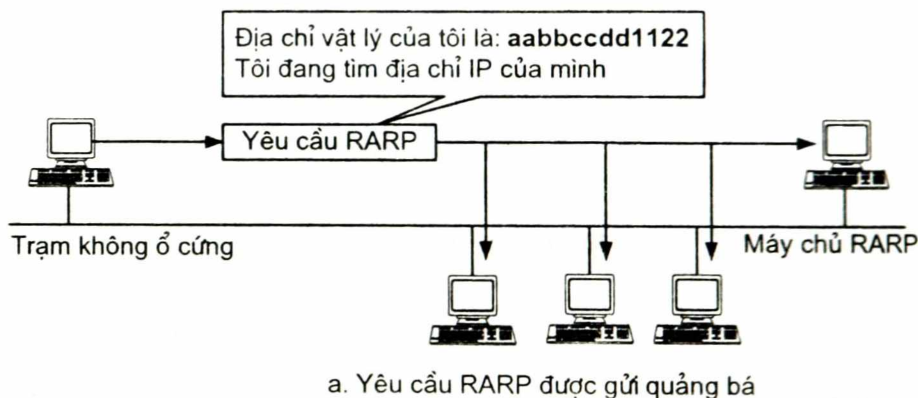
b. Trả lời ARP được gửi trực tiếp (unicast)

Hình 2.5: Hoạt động của ARP

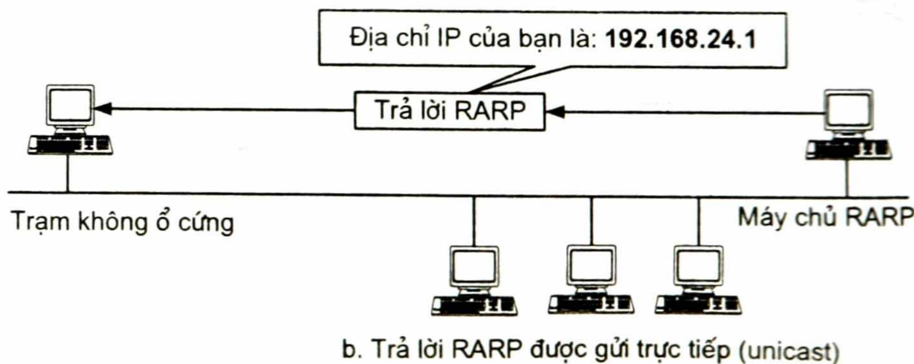
** Giao thức RARP*

RARP là giao thức chuyển đổi từ địa chỉ vật lý thành địa chỉ lô-gic. Nó được sử dụng trong trường hợp một máy biết địa chỉ vật lý của mình nhưng lại không biết địa chỉ IP, chẳng hạn một máy không có ổ cứng khởi động qua mạng.

Khi máy được bật, yêu cầu RARP được tạo ra và được gửi quảng bá trên mạng cục bộ. Một máy khác trên mạng biết về mọi địa chỉ IP sẽ trả lời yêu cầu bằng bản tin trả lời RARP. Máy yêu cầu RARP phải chạy chương trình RARP khách và máy trả lời RARP phải chạy chương trình RARP chủ.



a. Yêu cầu RARP được gửi quảng bá



b. Trả lời RARP được gửi trực tiếp (unicast)

Hình 2.6: Hoạt động của RARP

Trong hình 2.6, khi trạm không ổ cứng khởi động, yêu cầu RARP được tạo ra và được quảng bá tới mọi máy trên mạng. Mọi trạm trong mạng cục bộ đều nhận được gói này, nhưng chỉ máy chủ RARP trả lời yêu cầu. Gói trả lời có chứa địa chỉ IP của máy yêu cầu.

2.1.5 Tầng giao diện mạng

Tất cả mọi thứ nằm dưới IP được xem như một giao diện mạng. Tầng giao diện mạng gồm các giao thức thực hiện việc giao tiếp với mạng.

Có rất nhiều giao thức tầng vật lý và tầng liên kết dữ liệu có thể được sử dụng để truyền IP. Thực tế, một trong các mục đích của các nhà thiết kế IP là tạo cho nó tính linh hoạt cao, do vậy xuất hiện câu nói *IP trên tất cả.*

Tầng giao diện mạng còn được gọi là tầng truy nhập mạng. Các giao thức tầng giao diện mạng tiêu biểu có thể chạy dưới IP gồm:

- DIX Ethernet và 802.3 Ethernet, Fast Ethernet và Gigabit Ethernet
- Token Ring, cả loại 4 Mbit/s và 16 Mbit/s.
- FDDI
- SLIP (Serial Line IP)
- PPP (Point-to-Point Protocol)
- Chuyển tiếp khung (Frame Relay)
- ATM (Asynchronous Transfer Mode)
- SMDS (Switched Multimegabit Data Service)

2.2 ĐỊA CHỈ IP

Ở mức ứng dụng, chúng ta có thể coi một liên mạng là một mạng đơn lẻ kết nối các trạm với nhau. Để một trạm truyền thông với trạm khác, chúng ta cần một hệ thống định danh toàn cầu. Nói cách khác, chúng ta cần đặt tên duy nhất cho mỗi trạm.

Một liên mạng được tạo nên từ sự kết hợp của các mạng vật lý (LAN hoặc WAN) kết nối với nhau qua các bộ định tuyến. Khi một trạm truyền thông với một trạm khác, gói dữ liệu có thể di chuyển từ một mạng vật lý này đến mạng vật lý khác qua các bộ định tuyến. Vì thế, việc truyền thông tại mức này cũng cần có một hệ thống định danh toàn cục. Một trạm có thể truyền thông với một trạm bất kỳ mà không phải lo lắng về mạng vật lý phải đi qua. Nghĩa là tại tầng này, một trạm cũng phải được định danh duy nhất và toàn cục. Hơn nữa, để định tuyến tối ưu và hiệu quả, mỗi bộ định tuyến cũng phải được định danh duy nhất và toàn cục tại tầng này.

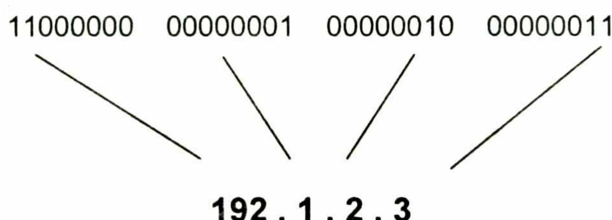
Số hiệu nhận dạng được sử dụng ở tầng liên mạng của bộ giao thức TCP/IP được gọi là địa chỉ liên mạng hay địa chỉ IP. Nó là địa chỉ nhị phân 32 bit (IPv4) hoặc 128 bit (IPv6), được thực hiện trong phần mềm, dùng để định danh duy nhất và toàn cục một trạm hoặc một bộ định tuyến trên liên mạng.

Các địa chỉ IP là duy nhất theo nghĩa mỗi địa chỉ định danh một và chỉ một thiết bị (máy trạm hoặc bộ định tuyến) trên liên mạng. Hai thiết bị trên liên mạng không thể có cùng địa chỉ IP. Tuy nhiên, một thiết bị có thể có nhiều địa chỉ IP nếu chúng được kết nối tới nhiều mạng vật lý khác nhau. Các địa chỉ IP là toàn cục theo nghĩa hệ thống đánh địa chỉ này phải được tất cả các trạm muốn kết nối tới liên mạng chấp nhận.

2.2.1 IPv4

Mỗi địa chỉ IPv4 gồm 4 byte (32 bit), định nghĩa hai phần: địa chỉ mạng (NetID) và địa chỉ trạm (HotID). Các phần này có chiều dài khác nhau tùy thuộc vào lớp địa chỉ. Các bit đầu tiên trong phần địa chỉ mạng xác định lớp của địa chỉ IP.

Để dễ đọc và dễ nhớ, các địa chỉ IP thường được biểu diễn dưới dạng thập phân dấu chấm. Trong cách biểu diễn này, các byte được tách riêng và được biểu diễn dưới dạng thập phân; phân tách các byte này là một dấu chấm (ví dụ hình 2.7).



Hình 2.7: Ký pháp thập phân dấu chấm

2.2.1.1 Các lớp địa chỉ IP

Địa chỉ IPv4 được chia làm 5 lớp, ký hiệu là A, B, C, D và E. Chiều dài phần địa chỉ mạng và phần địa chỉ trạm của các lớp là khác nhau. Cấu trúc của các lớp được chỉ ra trong hình 2.8.

Các bit đầu tiên của byte đầu tiên của địa chỉ IP được dùng để định danh lớp địa chỉ (0 - lớp A; 10 - lớp B; 110 - lớp C; 1110 - Lớp D và 1111 - lớp E).

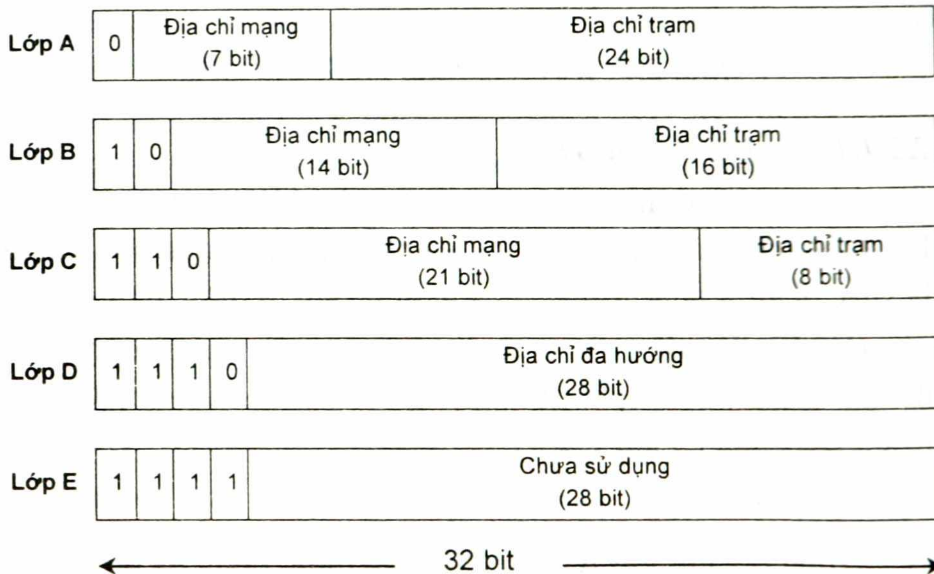
** Lớp A*

Trong địa chỉ lớp A, byte đầu tiên được dùng để định nghĩa địa chỉ mạng. Tuy nhiên bit đầu tiên phải luôn luôn bằng '0'. 7 bit còn lại định nghĩa các mạng khác nhau. Nghĩa là số mạng

có địa chỉ IP lớp A rất hạn chế. Về lý thuyết có thể có $2^7 = 128$ mạng lớp A. Tuy nhiên trên thực tế chỉ có 126 mạng vì có 2 mạng được dành riêng cho các mục đích cụ thể.

Trong một mạng chỉ lớp A, 24 bit được sử dụng để định danh địa chỉ trạm. Nghĩa là về lý thuyết có thể có tối đa $2^{24} = 16.777.216$ trạm. Tuy nhiên cũng có hai địa chỉ đặc biệt (phần địa chỉ trạm gồm toàn bit '0' hoặc toàn bit '1') được sử dụng làm các địa chỉ đặc biệt. Nghĩa là thực tế chỉ có tối đa 16.777.214 trạm trong một mạng lớp A.

Các địa chỉ lớp A được thiết kế cho các tổ chức có số lượng máy tính cực lớn kết nối vào mạng. Tuy nhiên, không một tổ chức nào có số lượng máy lớn như vậy do đó rất nhiều địa chỉ bị lãng phí.



Hình 2.8: Các lớp địa chỉ IP

** Lớp B*

Trong địa chỉ lớp B, 2 byte đầu được dùng để định nghĩa địa chỉ mạng và 2 byte sau để định nghĩa địa chỉ trạm. Tuy nhiên, hai bit đầu tiên trong phần địa chỉ mạng luôn luôn là '10', nên chỉ có 14 bit để định nghĩa các mạng khác nhau. Nghĩa là có nhiều mạng lớp B hơn lớp A. Số mạng lớp B là 2^{14} = 16.384.

Trong một mạng lớp B, 16 bit được sử dụng để định danh trạm, nghĩa là về lý thuyết mỗi mạng có thể có tối đa 2^{16} = 65536 trạm. Tuy nhiên cũng có hai địa chỉ đặc biệt nên thực tế một mạng lớp B chỉ có tối đa 65534 trạm.

Các địa chỉ lớp B được thiết kế cho các công ty cỡ vừa, những công ty có số lượng máy tính tương đối lớn. Tuy nhiên cũng giống những mạng lớp A, nhiều địa chỉ IP bị lãng phí vì rất ít công ty có số lượng máy tính lớn như vậy.

** Lớp C*

Trong địa chỉ lớp C, 3 byte đầu được dùng cho phần địa chỉ mạng và 1 byte cuối được dùng cho phần địa chỉ trạm. Tuy nhiên, 3 bit đầu tiên trong phần địa chỉ mạng luôn luôn là '110', nên chỉ còn 21 bit để định nghĩa địa chỉ mạng. Số mạng lớp C lớn hơn số mạng lớp A, B và bằng 2^{21} = 2.097.152 mạng.

Một mạng lớp C về lý thuyết có thể có tối đa 2^{8} = 256 trạm. Tuy nhiên thực tế chỉ có thể có tối đa 254 trạm do có hai địa chỉ được sử dụng cho các mục đích đặc biệt.

Địa chỉ lớp C được thiết kế cho các công ty nhỏ, những công ty chỉ có ít trạm nối vào mạng.

** Lớp D*

Địa chỉ lớp D được định nghĩa cho truyền đa hướng (multicasting). Trong lớp này, không có phần địa chỉ mạng và địa chỉ trạm. 4 bit đầu luôn luôn bằng '1110' để định nghĩa địa chỉ lớp D. 28 bit còn lại để định nghĩa địa chỉ đa hướng (multicast).

** Lớp E*

Lớp E được dự phòng để sử dụng cho các mục đích đặc biệt. Không có phần địa chỉ mạng và địa chỉ trạm. 4 bit đầu tiên bằng '1111' để định nghĩa lớp E.

2.2.1.2 Các địa chỉ đặc biệt

Một số địa chỉ trong khoảng địa chỉ lớp A, B và C được sử dụng cho các địa chỉ đặc biệt (bảng 2-1).

Bảng 2.1: Các địa chỉ đặc biệt

Địa chỉ đặc biệt	Phần địa chỉ mạng	Phần địa chỉ trạm	Nguồn hoặc đích
Địa chỉ mạng	Cụ thể	Toàn bit 0	Không dùng
Địa chỉ quảng bá trực tiếp	Cụ thể	Toàn bit 1	Đích
Địa chỉ quảng bá giới hạn	Toàn bit 1	Toàn bit 1	Đích
Địa chỉ lặp vòng	127	Bất kỳ	Đích

** Địa chỉ mạng*

Trong các lớp A, B và C, một địa chỉ có phần địa chỉ trạm gồm toàn bit 0 không được dùng cho bất cứ trạm nào. Nó được sử dụng để định nghĩa địa chỉ mạng. Nói cách khác, mạng được xem như một thực thể và có địa chỉ IP với phần địa chỉ trạm gồm

toàn bit '0'. Chú ý rằng địa chỉ mạng khác với phần địa chỉ mạng. Phần địa chỉ mạng chỉ là một phần của địa chỉ IP, còn địa chỉ mạng là một địa chỉ có phần địa chỉ trạm gồm toàn bit '0'. Địa chỉ này không thể sử dụng để định nghĩa một địa chỉ nguồn hoặc đích trong một gói IP. Ví dụ về địa chỉ mạng:

- Lớp A: 10.0.0.0
- Lớp B: 128.1.0.0
- Lớp C: 192.168.2.0

** Địa chỉ quảng bá trực tiếp (Direct Broadcast)*

Trong các địa chỉ lớp A, B và C, nếu phần địa chỉ trạm gồm toàn số '1' thì địa chỉ này được gọi là địa chỉ quảng bá trực tiếp. Địa chỉ này được bộ định tuyến sử dụng để gửi một gói tới tất cả các trạm trong một mạng cụ thể. Tất cả các trạm sẽ chấp nhận gói có loại địa chỉ này. Chú ý rằng địa chỉ này chỉ được sử dụng như địa chỉ đích trong một gói IP.

Ví dụ về địa chỉ quảng bá trực tiếp:

- Lớp A: 10.255.255.255
- Lớp B: 128.5.255.255
- Lớp C: 192.168.3.255

** Địa chỉ quảng bá giới hạn (Limited Broadcast)*

Nếu một địa chỉ có phần địa chỉ mạng gồm toàn bit '1' và địa chỉ trạm cũng gồm toàn bit '1' thì địa chỉ này được dùng để định nghĩa địa chỉ quảng bá trong mạng hiện tại. Một trạm muốn gửi một thông báo tới tất cả các trạm khác trên mạng có thể sử dụng địa chỉ này làm địa chỉ đích trong gói IP. Tuy nhiên bộ

định tuyến sẽ chặn các gói có địa chỉ loại này để hạn chế quảng bá trong mạng cục bộ. Chú ý rằng địa chỉ này (255.255.255.255) thuộc về lớp E. Tất cả các thiết bị trong mạng này đều nhận và xử lý gói tin.

* *Địa chỉ lặp vòng (loopback)*

Địa chỉ IP với byte đầu tiên là 127 được sử dụng làm địa chỉ lặp vòng, địa chỉ được sử dụng để kiểm tra phần mềm TCP/IP trên một máy. Khi địa chỉ này được sử dụng, gói sẽ không đi khỏi máy mà nó sẽ quay trở lại phần mềm giao thức. Địa chỉ này có thể được sử dụng để kiểm tra phần mềm IP. Ví dụ, một ứng dụng, chẳng hạn "Ping" có thể gửi một gói với địa chỉ đích là địa chỉ lặp vòng để kiểm tra xem phần mềm IP có khả năng nhận và xử lý gói hay không.

Một ví dụ khác là địa chỉ lặp vòng có thể sử dụng bởi một tiến trình khách (một ứng dụng đang chạy) để gửi một thông báo tới một tiến trình chủ trên cùng một máy. Chú ý rằng địa chỉ lặp vòng chỉ được sử dụng như địa chỉ đích trong một gói IP.

2.2.1.3 Địa chỉ riêng

Trong một mạng biệt lập (không nối tới Internet), người quản trị có thể sử dụng bất kỳ địa chỉ nào mình muốn. Tuy nhiên, để tránh sự nhầm lẫn giữa một địa chỉ thực trên Internet và một địa chỉ dùng trong một mạng riêng, tổ chức cấp số Internet đã dành một số khối địa chỉ để sử dụng cho mạng riêng. Các khối địa chỉ này không được cấp cho các mạng tham gia vào Internet.

Các địa chỉ dùng cho mạng riêng như sau:

Lớp A: 10.0.0.0 (1 mạng)

Lớp B: 172.16.0.0 đến 172.31.0.0 (16 mạng)

Lớp C: 192.168.0.0 đến 192.168.255.0 (256 mạng).

2.2.1.4 Mạng con và siêu mạng

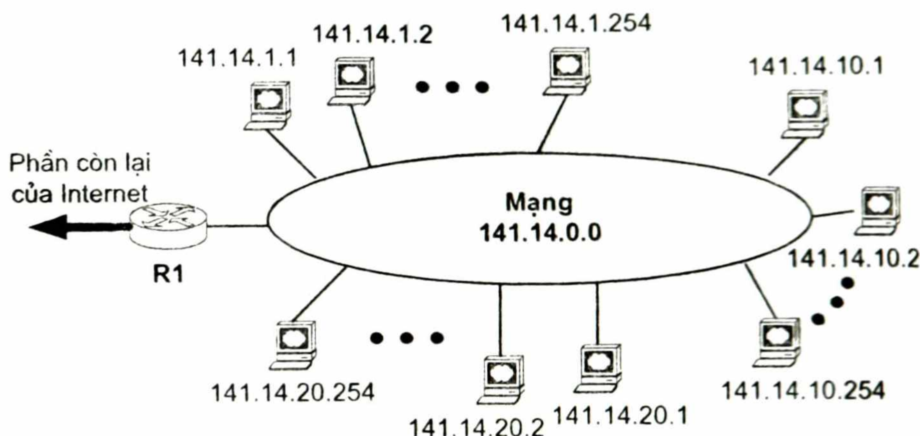
Phần này trình bày về hai vấn đề liên quan đến đánh địa chỉ IP: phân mạng con và lập siêu mạng. Phân mạng con là quá trình chia mạng thành các mạng nhỏ hơn (mạng con); mỗi mạng có một địa chỉ riêng. Lập siêu mạng là quá trình kết hợp nhiều mạng để tạo thành một mạng lớn hơn (siêu mạng).

** Mạng con (subnet)*

Ở phần trên chúng ta đã biết địa chỉ IP gồm hai phần: địa chỉ mạng (NetID) và địa chỉ trạm (HostID). Nghĩa là ta có một lược đồ đánh địa chỉ phân cấp. Để tới được một trạm trong liên mạng, trước tiên ta phải tới được mạng chứa trạm đó bằng cách sử dụng địa chỉ mạng. Sau đó ta phải sử dụng địa chỉ trạm để tới được trạm trong mạng. Nói cách khác, các lớp địa chỉ A, B, C được thiết kế với hai mức phân cấp (hình 2.9).

Với lược đồ này, mạng của một công ty bị hạn chế ở hai mức phân cấp. Công ty không thể tổ chức các trạm thành nhóm mà mọi trạm đều ở cùng một mức. Do đó, công ty có một mạng với rất nhiều máy tính.

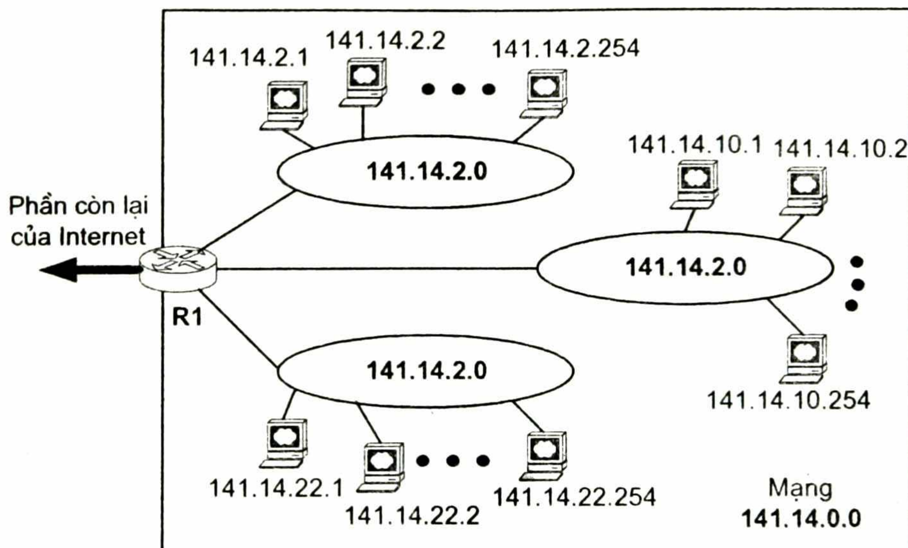
Một giải pháp cho vấn đề này là phân mạng con. Nghĩa là chia mạng thành các mạng nhỏ hơn, được gọi là mạng con.



Hình 2.9: Mạng với hai mức phân cấp (chưa phân mạng con)

Trong ví dụ ở trên, phần còn lại của Internet không biết rằng mạng được chia thành ba mạng vật lý nhỏ hơn (hình 2.10). Đối với phần còn lại của Internet, ba mạng con này vẫn như là một mạng. Một gói gửi tới máy 141.14.2.21 vẫn tới bộ định tuyến R1. Địa chỉ đích của gói dữ liệu IP vẫn là địa chỉ lớp B với địa chỉ mạng là 141.14 và địa chỉ trạm là 2.21.

Tuy nhiên, khi gói dữ liệu tới bộ định tuyến R1, sự biên dịch địa chỉ IP thay đổi. Bộ định tuyến R1 biết rằng mạng 141.14 được chia thành ba mạng con. Nó biết rằng hai byte sau trong địa chỉ IP xác định hai thứ: phần địa chỉ mạng con (SubnetID) và phần địa chỉ trạm (HostID). Do đó, 2.21 phải được hiểu là địa chỉ mạng con 2 và địa chỉ trạm 21. Bộ định tuyến R1 sử dụng hai byte đầu là địa chỉ mạng, byte thứ ba là địa chỉ mạng con và byte cuối cùng là địa chỉ trạm.



Hình 2.10: Mạng với ba mức phân cấp (phân mạng con)

Thêm các mạng con tạo ra một mức phân cấp trung gian trong hệ thống đánh địa chỉ IP. Bây giờ ta có ba mức: địa chỉ mạng; địa chỉ mạng con và địa chỉ trạm. Địa chỉ mạng là mức đầu tiên, nó xác định mạng. Mức thứ hai là địa chỉ mạng con; nó xác định mạng con. Địa chỉ trạm là mức thứ ba; nó xác định trạm trong mạng con.

Quá trình chuyển gói được chia thành ba bước: chuyển gói tới mạng; chuyển gói tới mạng con; chuyển gói tới trạm.

Mặt nạ mạng con

Mặt nạ mạng con không phải là một địa chỉ, nhưng nó xác định đâu là phần địa chỉ mạng và đâu là phần địa chỉ trạm trong một địa chỉ IP. Mặt nạ mạng con cũng dài 32 bit giống địa chỉ IP. Nó gồm hai phần: phần đầu gồm toàn bit 1 cho biết chiều dài

phần địa chỉ mạng, phần cuối gồm toàn bit 0 cho biết phần địa chỉ trạm.

Thực hiện các bước sau để xác định mặt nạ mạng con cho một địa chỉ IP mạng con cụ thể:

Bước 1: Biểu diễn địa chỉ IP mạng con dưới dạng nhị phân

Bước 2: Thay phần địa chỉ mạng và địa chỉ mạng con trong địa chỉ bằng các bit 1

Bước 3: Thay phần địa chỉ trạm bằng các bit 0

Bước 4: Chuyển biểu diễn nhị phân về dạng thập phân.

Mặt nạ mạng con luôn đi kèm với địa chỉ IP phân mạng con; đối với các địa chỉ IP không phân mạng con, ta dùng mặt nạ mạng con mặc định. Mặt nạ mạng con mặc định đối với địa chỉ lớp A là 255.0.0.0; lớp B là 255.255.0.0 và lớp C là 255.255.255.0.

Phân mạng con

Như ta đã biết, lớp A, lớp B, lớp C dành 8, 16, 24 bit cho phần địa chỉ mạng. Để phân mạng con, bạn phải mở rộng địa chỉ mạng, nghĩa là thêm địa chỉ mạng con. Phần địa chỉ mạng con được tạo ra bằng cách mượn một số bit trong phần địa chỉ trạm.

Khi phân mạng con, ta không được phép sử dụng các mạng con có phần địa chỉ mạng con gồm toàn bit 0 hoặc toàn bit 1. Nếu ta mượn 1 bit thì có thể chia mạng thành hai mạng con, nhưng đều không thể sử dụng. Nếu ta mượn tất cả các bit trong phần địa chỉ trạm thì không còn phần địa chỉ trạm nữa. Nếu ta mượn toàn bộ, chỉ trừ 1 bit cho phần địa chỉ trạm thì số trạm có thể sử dụng trong mỗi mạng con bằng 0 vì ta không được dùng các địa chỉ có phần địa chỉ trạm gồm toàn bit 1 hoặc toàn bit 0.

Do đó, số bit ta có thể mượn khi phân mạng con phải lớn hơn hoặc bằng 2 và nhỏ hơn hoặc bằng n-2, với n là số bit trong phần địa chỉ trạm.

Nếu ta mượn m bit thì số lượng mạng con thu được là 2^m, và số lượng mạng con có thể sử dụng là 2^m-2.

Mặt nạ mạng con chiều dài biến thiên (VLSM)

Mặc dù việc phân mạng con là một bổ sung có giá trị cho kiến trúc đánh địa chỉ IP, nhưng nó vẫn có một hạn chế cơ bản. Đó là bạn phải sử dụng một mặt nạ mạng con cho toàn bộ mạng. Do đó, sau khi bạn đã chọn một mặt nạ mạng con, bạn không thể cung cấp các mạng con có kích thước khác. Nếu có yêu cầu tăng kích thước của mạng con, bạn phải thay đổi lại mặt nạ mạng con cho toàn bộ mạng. Đây là một công việc phức tạp và tốn thời gian.

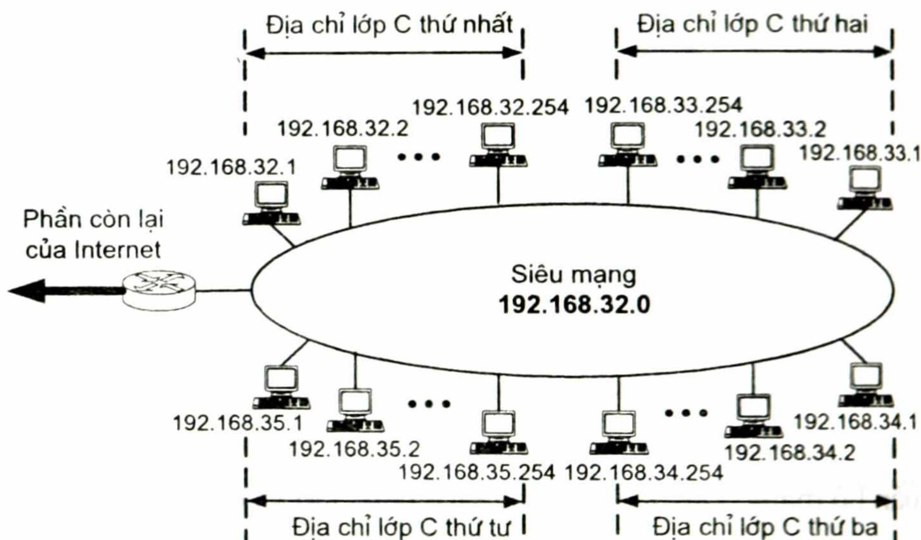
Một giải pháp cho vấn đề này là cho phép một mạng được phân mạng con có thể có nhiều mặt nạ mạng con. Đương nhiên, mỗi mặt nạ mạng con có một kích thước khác nhau. Kỹ thuật mới về phân mạng con này được gọi là mặt nạ mạng con chiều dài biến thiên (VLSM – Variable Length Subnet Masking).

** Siêu mạng (Supernet)*

Trong khi các địa chỉ lớp A và lớp B gần như đã cạn kiệt, thì địa chỉ lớp C vẫn còn. Tuy nhiên, khoảng địa chỉ lớp C, với tối đa 254 trạm, có thể không thỏa mãn nhu cầu của các công ty có số lượng máy tính lớn hơn 254 máy.

Giải pháp cho vấn đề này chính là lập siêu mạng. Để làm vậy, một công ty có thể sử dụng một khối địa chỉ lớp C thay vì chỉ

sử dụng một. Ví dụ, nếu công ty cần 1000 địa chỉ, thì họ có thể sử dụng 4 địa chỉ lớp C trong một siêu mạng. Hình 2.11 minh họa việc kết hợp 4 địa chỉ lớp C để tạo thành một siêu mạng.



Hình 2.11: Siêu mạng

Mặt nạ siêu mạng

Một mặt nạ siêu mạng có thể được gán cho một khối địa chỉ mạng lớn C nếu số lượng địa chỉ mạng trong khối đó bằng 2^n (2, 4, 8, 16, ...). Mặt nạ mặc định của địa chỉ lớp C là 255.255.255.0, 24 số 1 theo sau bởi 8 số 0. Nếu một số bit 1 được chuyển thành 0, ta có thể có mặt nạ cho một nhóm địa chỉ mạng lớp C. Như vậy, mặt nạ siêu mạng ngược với mặt nạ mạng con. Trong mặt nạ mạng con, ta thay đổi một số bit 0 trong phần địa chỉ trạm thành 1. Trong mặt nạ siêu mạng, ta thay đổi một số bit 1 trong phần địa chỉ mạng thành 0.

Khi biểu diễn một siêu mạng, ta dùng địa chỉ IP thấp nhất và mặt nạ siêu mạng. Chẳng hạn, siêu mạng ở hình 2.11 được

biểu diễn là: 192.168.32.0 255.255.252.0 chứ không phải là 192.168.33.0 255.255.252.0.

Ví dụ, với mặt nạ siêu mạng 255.255.255.252, ta có thể có 4 địa chỉ lớp C kết hợp trong một siêu mạng. Nếu chúng ta chọn địa chỉ đầu tiên là 192.168.32.0 thì ba địa chỉ khác là 192.168.33.0; 192.168.34.0; 192.168.35.0. Khi bộ định tuyến nhận được một gói dữ liệu, nó áp dụng mặt nạ siêu mạng cho địa chỉ đích trong gói và so sánh kết quả với địa chỉ thấp nhất. Nếu kết quả và địa chỉ thấp nhất giống nhau, gói thuộc về siêu mạng (gửi tới siêu mạng).

Giả sử gói đến có địa chỉ đích là 192.168.34.4. Sau khi áp dụng mặt nạ siêu mạng, kết quả là 192.168.32.0, nghĩa là giống địa chỉ thấp nhất. Do vậy, gói thuộc về siêu mạng.

Bây giờ, giả sử gói đến có địa chỉ đích là 192.168.37.4. Sau khi áp dụng mặt nạ siêu mạng, kết quả là 192.168.36.0, nghĩa là không giống địa chỉ thấp nhất. Do vậy, gói không thuộc về siêu mạng.

Định tuyến liên vùng không phân lớp (CIDR)

Lập siêu mạng có nghĩa gán một tập địa chỉ lớp C cho một công ty cần nhiều hơn 254 địa chỉ trạm. Tuy nhiên, khi những địa chỉ lớp C này được đưa vào bảng định tuyến, mỗi địa chỉ sẽ chiếm một mục trong bảng định tuyến. Điều này nghĩa là nếu một công ty dùng 256 địa chỉ lớp C (thay vì một địa chỉ lớp B), thì sẽ có 256 mục trong bảng định tuyến. Kỹ thuật định tuyến liên vùng không liên kết (CIDR) được phát minh để giảm số mục trong bảng định tuyến.

Trong kỹ thuật này, thay vì đưa mỗi địa chỉ lớp C cùng mặt nạ mạng con mặc định tương ứng vào bảng định tuyến, bộ định tuyến có thể sử dụng mặt nạ siêu mạng địa chỉ mạng thấp nhất trong nhóm.

2.2.2 IPv6

Giao thức tầng liên mạng trong bộ giao thức TCP/IP hiện nay là IPv4. IPv4 cung cấp truyền thông trạm-tới-trạm giữa các hệ thống trên Internet. Cho dù IPv4 được thiết kế tốt, nhưng nó vẫn có một số hạn chế mà trở nên không phù hợp với sự phát triển nhanh của Internet:

- IPv4 sử dụng 32 bit để đánh địa chỉ và chia khoảng địa chỉ thành các lớp A, B, C, D, E. Khoảng địa chỉ IPv4 là không đủ cho sự phát triển nhanh chóng của Internet.

- Internet phải điều tiết truyền video và âm thanh thời gian thực. Các kiểu truyền này yêu cầu độ trễ tối thiểu và khả năng dành trước tài nguyên. Những điều chưa được hỗ trợ trong IPv4.

- IPv4 không hỗ trợ mật mã và chứng thực.

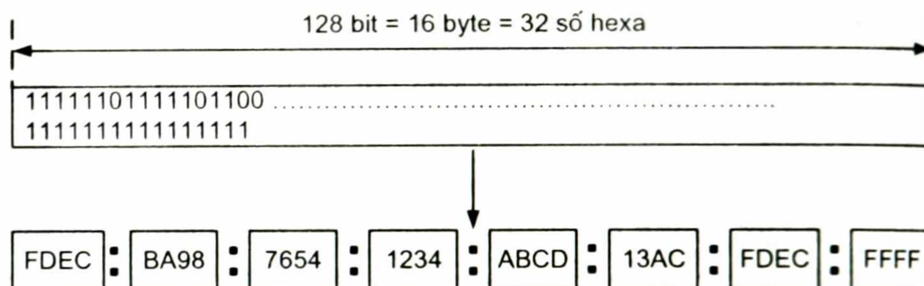
Để vượt qua những hạn chế này, IPv6 đã được phát triển. Trong IPv6, giao thức IP được thay đổi để điều tiết sự thay đổi không thể đoán trước của Internet. Định dạng và chiều dài của địa chỉ IP được thay đổi cùng với định dạng gói. Các giao thức liên quan như ICMP cũng được thay đổi. Các giao thức khác ở tầng liên mạng như ARP, RARP và IGMP không còn. Chức năng của chúng được đưa vào giao thức ICMP. Các giao thức định tuyến, như RIP và OSPF, cũng có chút sửa đổi để thích ứng những thay đổi này.

IPv6 có một số ưu điểm so với IPv4:

- Khoảng địa chỉ lớn hơn: Địa chỉ IPv6 dài 128 bit, nghĩa là gấp bốn lần chiều dài địa chỉ IPv4.
- Định dạng tiêu đề tốt hơn: IPv6 sử dụng định dạng tiêu đề mới, trong đó các tùy chọn được tách khỏi phần tiêu đề cơ sở và nếu cần, được thêm vào giữa phần tiêu đề cơ sở và dữ liệu. Do vậy, làm đơn giản và tăng tốc độ xử lý định tuyến vì hầu hết các tùy chọn đều không cần được bộ định tuyến kiểm tra.
- Các tùy chọn mới: IPv6 có một số tùy chọn mới cho phép các chức năng bổ sung.
- Cho phép mở rộng: IPv6 được thiết kế để cho phép mở rộng khi có yêu cầu.
- Hỗ trợ cấp phát tài nguyên: Trong IPv6, trường loại dịch vụ được bỏ đi, nhưng một cơ chế (được gọi là nhãn luồng) được thêm vào để cho phép nguồn yêu cầu xử lý gói đặc biệt. Cơ chế này có thể được sử dụng để hỗ trợ lưu lượng video hoặc âm thanh thời gian thực.
- Bảo mật hơn: Tùy chọn mật mã và chứng thực trong IPv6 cung cấp tính toàn vẹn và tính bảo mật của gói dữ liệu.

2.2.2.1 Địa chỉ IPv6

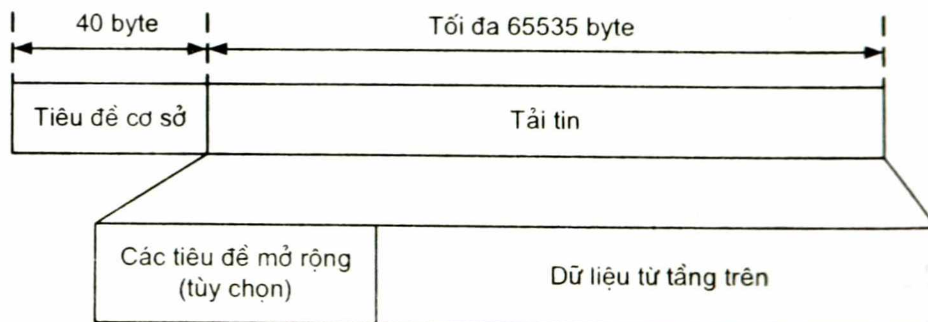
Địa chỉ IPv6 dài 128 bit và được biểu diễn dưới dạng số hexa hai chấm. Trong cách biểu diễn này, 128 bit được chia thành 8 phần, mỗi phần dài 2 byte. Hai byte được biểu diễn bằng 4 số hexa. Do đó, địa chỉ IP gồm 32 số hexa, cứ 4 số hexa có một dấu hai chấm để phân tách (xem ví dụ hình 2.12).



Hình 2.12: Ví dụ về địa chỉ IPv6

2.2.2.2 Định dạng gói IPv6

Định dạng gói IPv6 được cho ở hình 2.13. Mỗi gói gồm hai phần: tiêu đề cơ sở và tải tin. Tải tin lại gồm hai phần: các tiêu đề mở rộng tùy chọn và dữ liệu tầng trên. Tiêu đề cơ sở chiếm 40 byte, trong khi tiêu đề mở rộng và dữ liệu có chiều dài tối đa 65535 byte.



Hình 2.13: Gói IPv6

* *Tiêu đề cơ sở*

Hình 2.14 minh họa định dạng gói IPv6.

<table>
<tr><td>Phiên bản</td><td>Lớp lưu lượng</td><td colspan="2">Nhãn luồng</td></tr>
<tr><td colspan="2">Độ dài tải tin</td><td>Tiêu đề tiếp theo</td><td>Giới hạn bước nhảy</td></tr>
<tr><td colspan="4">Địa chỉ nguồn</td></tr>
<tr><td colspan="4">Địa chỉ đích</td></tr>
<tr><td colspan="4">Tải tin
(Các tiêu đề mở rộng + Dữ liệu từ tầng trên)</td></tr>
</table>

Hình 2.14: Định dạng gói IPv6

Các trường trong phần tiêu đề cơ sở như sau:

- *Phiên bản (Version):* Trường 4 bit này cho biết phiên bản của gói IP. Đối với IPv6, giá trị trường này là 6.

- *Lớp lưu lượng (Traffic Class):* Trường 8 bit này định nghĩa độ ưu tiên hoặc lớp của gói tương tự như trường ToS trong IPv4, nó có thể sử dụng trong trường hợp có tắc nghẽn.

- *Nhãn luồng (Flow label):* Trường 20 bit này được thiết kế để cung cấp sự xử lý đặc biệt cho một luồng dữ liệu cụ thể. Chúng ta sẽ thảo luận trường này ở phần sau.

- *Độ dài tải tin (payload length):* Trường hai byte này định nghĩa độ dài của gói IP, không tính phần tiêu đề cơ sở.

- *Tiêu đề tiếp theo (next header):* Trường 8 bit này định nghĩa tiêu đề theo sau phần tiêu đề cơ sở. Tiêu đề tiếp theo có

thể là một tiêu đề mở rộng tùy chọn được IP sử dụng hoặc tiêu đề cho một giao thức tầng trên, chẳng hạn UDP hoặc TCP. Mỗi tiêu đề mở rộng đều chứa trường này.

- *Giới hạn bước nhảy (hop limit):* Trường 8 bit này có tác dụng tương tự như trường TTL trong IPv4.

- *Địa chỉ nguồn (source address):* Trường 128 bit này nhận dạng địa chỉ nguồn của gói tin.

- *Địa chỉ đích (Destination Address).* Trường 128 bit này thường nhận dạng đích cuối cùng của gói tin. Tuy nhiên, nếu định tuyến nguồn được sử dụng, trường này chứa địa chỉ của bộ định tuyến kế tiếp.

* *Nhãn luồng (Flow Label)*

Một dãy các gói được gửi từ một nguồn cụ thể tới một đích cụ thể mà cần được bộ định tuyến xử lý đặc biệt thì được gọi là một luồng gói. Sự kết hợp giữa địa chỉ IP nguồn và giá trị nhãn luồng định nghĩa duy nhất một luồng gói.

Đối với bộ định tuyến, luồng là một dãy gói chia sẻ cùng các đặc tính, chẳng hạn đi trên cùng một đường đi, sử dụng cùng tài nguyên, có loại bảo mật như nhau v.v...Một bộ định tuyến hỗ trợ xử lý nhãn luồng có một bảng nhãn luồng. Bảng này có mỗi mục cho mỗi nhãn luồng, mỗi mục định nghĩa các dịch vụ nhãn luồng tương ứng yêu cầu. Khi bộ định tuyến nhận gói, nó tham khảo bảng nhãn luồng để tìm mục tương ứng với giá trị nhãn luồng định nghĩa trong gói. Sau đó nó cung cấp cho gói các dịch vụ được đề cập trong mục này. Tuy nhiên, lưu ý rằng, nhãn luồng tự nó không cung cấp thông tin cho các mục trong bảng

nhãn luồng; thông tin được cung cấp bằng các phương tiện khác như các tùy chọn bước nhảy-bước nhảy hoặc các giao thức khác.

Ở dạng đơn giản nhất, nhãn luồng có thể được sử dụng để tăng tốc độ xử lý gói bởi bộ định tuyến. Khi bộ định tuyến nhận một gói, thay vì tham khảo đến bảng định tuyến và thông qua một giải thuật định tuyến để xác định địa chỉ của bước nhảy tiếp theo, nó có thể dễ dàng tìm bước nhảy tiếp theo bằng cách tìm trong bảng nhãn luồng.

Ở dạng phức tạp hơn, nhãn luồng có thể được sử dụng để hỗ trợ truyền video và âm thanh thời gian thực. Loại lưu lượng này thường yêu cầu các tài nguyên như băng thông cao, bộ đệm lớn, thời gian xử lý dài v.v... Một tiến trình có thể dành trước những tài nguyên này để đảm bảo rằng dữ liệu thời gian thực sẽ không bị trễ do thiếu tài nguyên. Để sử dụng dữ liệu thời gian thực và dành trước tài nguyên, cần bổ sung giao thức thời gian thực (RTP) và giao thức dành trước tài nguyên (RSVP) vào IPv6.

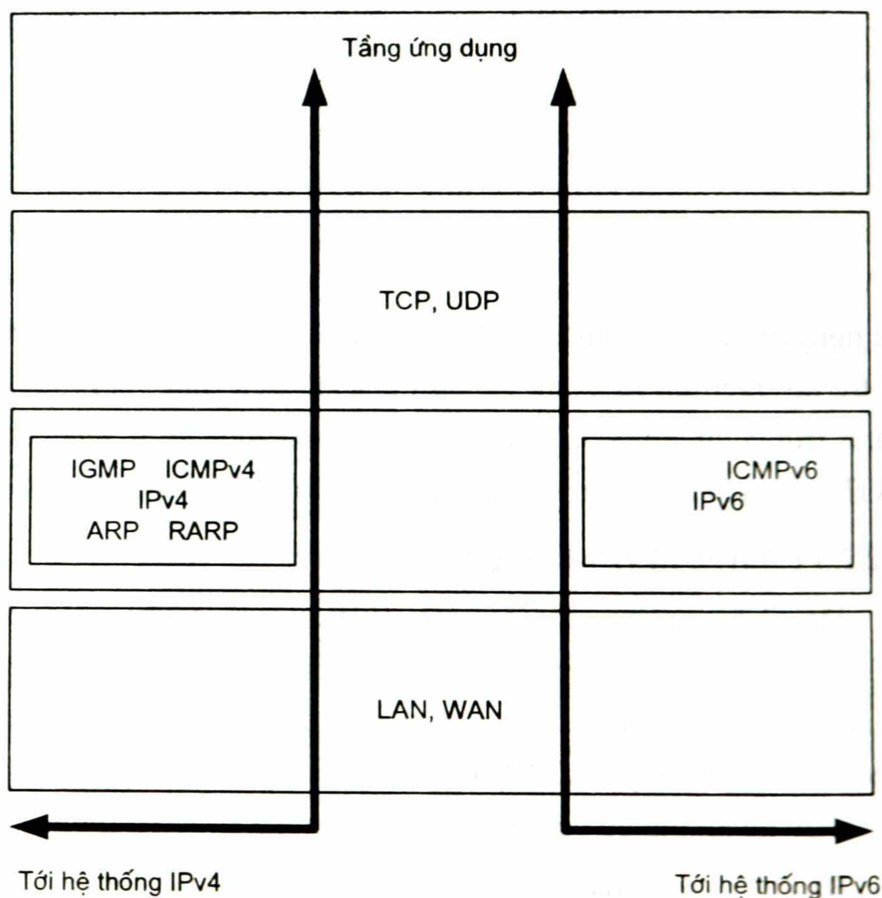
2.2.2.3 Chuyển từ IPv4 sang IPv6

Do số lượng nút trên Internet là một con số khổng lồ, nên việc chuyển dịch từ IPv4 sang IPv6 không thể xảy ra đột ngột. Phải mất một khoảng thời gian đáng kể để các hệ thống trên Internet có thể chuyển từ IPv4 sang IPv6. Việc chuyển dịch phải trôi chảy để không có vấn đề gì giữa hệ thống IPv4 và IPv6.

Ba chiến lược đã được IETF phát triển để sự chuyển dịch trôi chảy hơn, gồm: chồng giao thức kép, đường hầm và dịch tiêu đề.

** Chồng giao thức kép*

Chiến lược này khuyến nghị rằng mọi trạm trước khi chuyển đổi hoàn toàn sang IPv6, phải có một chồng giao thức kép. Nói cách khác, một trạm phải chạy IPv4 và IPv6 đồng thời cho đến khi toàn bộ Internet đều sử dụng IPv6. Hình 2.15 cho thấy một cấu hình chồng giao thức kép tổng quan.



Hình 2.15: Cấu hình chồng giao thức kép

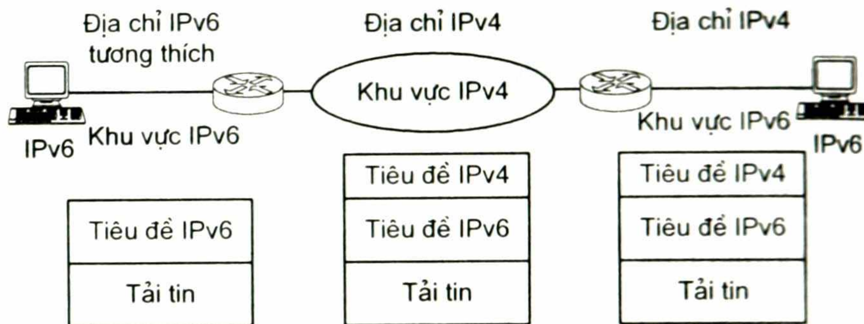
Để xác định sử dụng phiên bản nào khi gửi một gói tới đích, trạm nguồn truy vấn DNS. Nếu DNS trả lại địa chỉ IPv4, trạm nguồn gửi gói IPv4. Nếu DNS trả lại địa chỉ IPv6, trạm nguồn gửi gói IPv6.

**** Đường hầm***

Đường hầm là chiến lược được sử dụng khi hai máy tính IPv6 muốn truyền thông với nhau, nhưng gói phải đi qua một khu vực sử dụng IPv4. Để qua khu vực này, gói phải có một địa chỉ IPv4. Gói IPv6 được đóng gói trong gói IPv4 khi nó đi vào khu vực và được mở gói khi nó rời khỏi khu vực. Dường như gói IPv6 đi vào đường hầm ở một đầu và ló ra ở đầu kia.

Đường hầm tự động

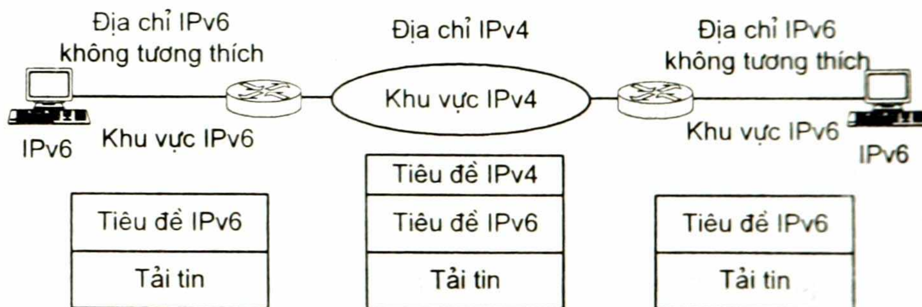
Nếu trạm nhận sử dụng địa chỉ tương thích IPv6, đường hầm xảy ra tự động mà không cần cấu hình lại. Máy gửi gửi cho máy nhận một gói IPv6 sử dụng địa chỉ tương thích IPv6. Khi gói đến đường biên của mạng IPv4, bộ định tuyến đóng gói nó trong gói IPv4 và đương nhiên nó cần địa chỉ IPv4. Để lấy địa chỉ này, bộ định tuyến trích địa chỉ IPv4 được nhúng trong địa chỉ IPv6. Sau đó, gói hoàn thành chuyến đi của mình như là một gói IPv4. Trạm đích, chạy chồng giao thức kép, nhận gói IPv4. Nó nhận ra địa chỉ IPv4 của nó, đọc phần tiêu đề và thấy rằng (thông qua giá trị trường giao thức) gói này đang mang gói IPv6. Sau đó, nó chuyển gói tới phần mềm IPv6 để xử lý.



Hình 2.16: Đường hầm tự động

Đường hầm cấu hình

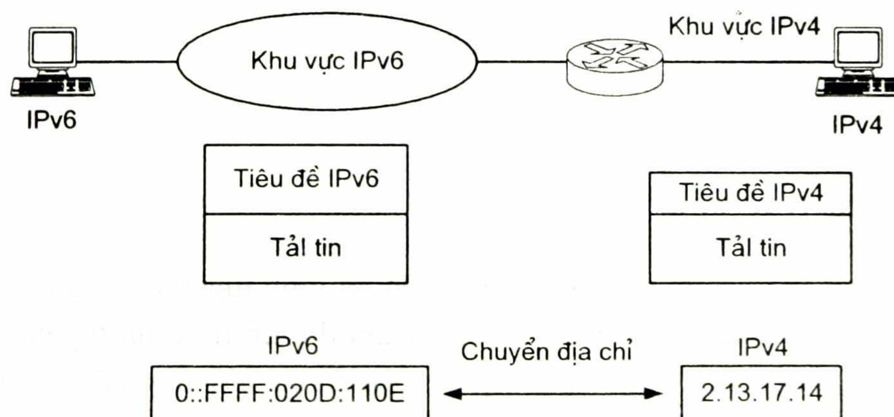
Nếu trạm nhận không hỗ trợ một địa chỉ tương thích IPv6, máy gửi nhận được một địa chỉ IPv6 không tương thích từ DNS. Trong trường hợp này, máy gửi gửi gói IPv6 với địa chỉ IPv6 không tương thích của máy nhận. Tuy nhiên, gói không thể đi qua khu vực IPv4 nếu không được đóng gói trong một gói IPv4. Hai bộ định tuyến tại biên của khu vực IPv4 được cấu hình để chuyển gói đã được đóng gói trong gói IPv4. Bộ định tuyến phía này gửi gói IPv4 với địa chỉ nguồn là địa chỉ IPv4 của nó và địa chỉ đích là địa chỉ IPv4 của bộ định tuyến kia. Bộ định tuyến kia nhận gói, mở gói để lấy gói IPv6 và chuyển gói tới đích.



Hình 2.17: Đường hầm cấu hình

** Dịch tiêu đề*

Chiến lược dịch tiêu đề cần thiết khi phần lớn Internet đã chuyển sang IPv6 và chỉ còn một phần nhỏ sử dụng IPv4. Máy gửi muốn sử dụng IPv6, nhưng trạm nhận không hiểu IPv6. Đường hầm không hoạt động trong trường hợp này vì gói phải ở định dạng IPv4 để trạm nhận có thể hiểu được. Trong trường hợp này, định dạng tiêu đề phải được thay đổi toàn bộ thông qua quá trình dịch tiêu đề. Tiều đề IPv6 được chuyển đổi thành tiêu đề IPv4 (hình 2.18).



Hình 2.18: Dịch tiêu đề

Dịch địa chỉ sử dụng địa chỉ được ánh xạ để dịch một địa chỉ IPv6 thành địa chỉ IPv4. Dưới đây là một số luật sử dụng để dịch tiêu đề gói IPv6 sang IPv4.

- Địa chỉ ánh xạ IPv6 được chuyển thành địa chỉ IPv4 bằng cách trích 32 bit phía bên phải.
- Giá trị trường độ ưu tiên của IPv6 được bỏ đi.
- Đặt trường loại dịch vụ trong IPv4 là 0.

- Tính mã tổng kiểm tra cho gói IPv4 và điền vào trường tương ứng.
- Bỏ qua nhãn luồng của IPv6.
- Các tiêu đề mở rộng tương thích được chuyển thành các tùy chọn và được đưa vào phần tiêu đề IPv4.
- Độ dài tiêu đề IPv4 được tính và được đưa vào trường tương ứng.
- Độ dài tổng của gói IPv4 được tính và được đưa vào trường tương ứng.

2.3. ĐỊNH TUYẾN

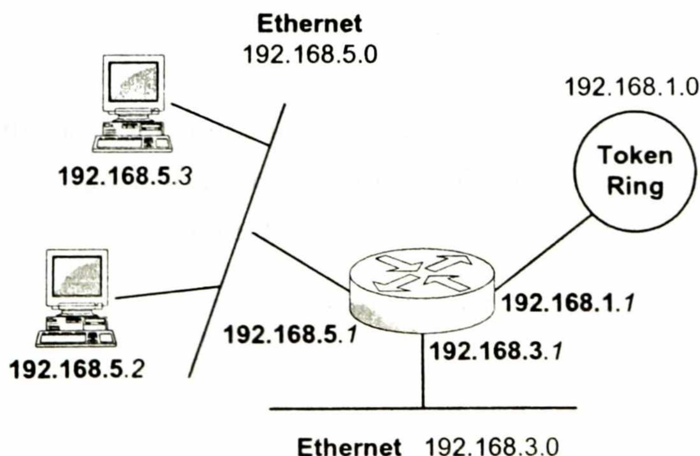
2.3.1 Khái niệm định tuyến

Định tuyến là phương thức di chuyển thông tin trong liên mạng, từ nguồn đến đích. Nó là một chức năng được thực hiện ở tầng mạng. Chức năng này cho phép bộ định tuyến đánh giá các đường đi sẵn có tới đích. Để đánh giá đường đi, định tuyến sử dụng các thông tin tô-pô mạng. Các thông tin này có thể do người quản trị thiết lập hoặc được thu thập thông qua các giao thức định tuyến.

Tầng mạng hỗ trợ chuyển gói đầu cuối-tới-đầu cuối theo kiểu cố gắng tối đa (best-effort) qua các mạng được kết nối với nhau. Tầng mạng sử dụng bảng định tuyến IP để gửi các gói từ mạng nguồn đến mạng đích. Sau khi đã quyết định sử dụng đường đi nào, bộ định tuyến tiến hành việc chuyển gói. Nó lấy một gói nhận được ở giao diện vào và chuyển tiếp gói này tới giao diện ra tương ứng (giao diện thể hiện đường đi tốt nhất tới đích cho gói).

Trong một liên mạng, mỗi mạng được định danh bởi một địa chỉ mạng và bộ định tuyến sử dụng các địa chỉ mạng này để nhận biết đích. Bộ định tuyến sử dụng địa chỉ mạng để nhận dạng mạng đích (LAN) của một gói tin trong một liên mạng. Hình 2.19 minh họa ba địa chỉ mạng được dùng để nhận diện các phân đoạn kết nối tới bộ định tuyến.

Mạng	**Trạm**
192.168.1	1
192.168.3	1
192.168.5	1 2 3



Hình 2.19: Một địa chỉ gồm phần địa chỉ mạng và địa chỉ trạm

** Các hành động trong quá trình định tuyến*

Khi định tuyến dữ liệu từ nguồn đến đích, bộ định tuyến thường chuyển tiếp gói từ một liên kết dữ liệu (mạng) này đến một liên kết dữ liệu khác, sử dụng hai chức năng cơ bản:

- Xác định đường đi (*path determination*)

- Chuyển mạch (*switching*)

Chức năng xác định đường đi chọn ra một đường đi tối ưu đến đích theo một tiêu chí nào đó (chẳng hạn chiều dài đường đi). Để trợ giúp cho quá trình xác định đường đi, các giải thuật định tuyến được khởi tạo và duy trì trong bảng định tuyến, bảng này chứa thông tin về các tuyến tới đích. Khi đường đi tối ưu được xác định, bước nhảy tiếp theo gắn với đường đi này cho bộ định tuyến biết phải gửi gói đi đâu để nó có thể đến đích theo đường đi tối ưu đó.

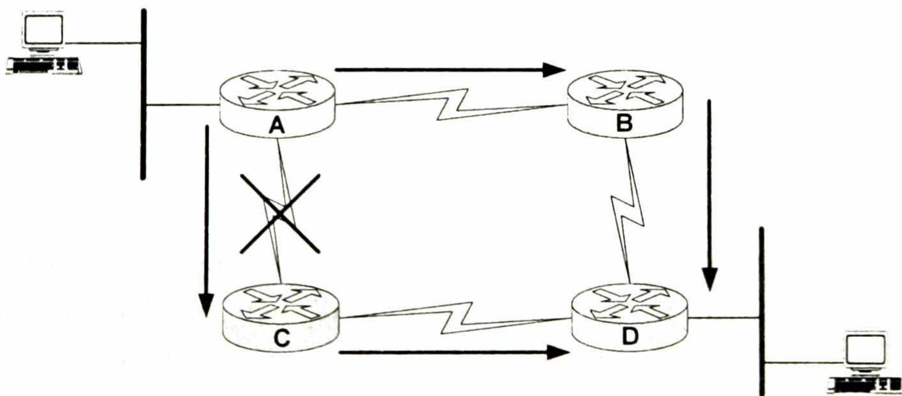
Chức năng chuyển mạch cho phép bộ định tuyến chuyển gói từ cổng vào tới cổng ra tương ứng với đường đi tối ưu đã chọn. Trong quá trình định tuyến, phần địa chỉ mạng được sử dụng để xác định đường đi, còn phần địa chỉ trạm được bộ định tuyến cuối cùng trên đường đi (bộ định tuyến nối trực tiếp tới mạng đích) sử dụng để chuyển gói tới đúng trạm đích.

** Định tuyến tĩnh và định tuyến động*

Các tuyến tĩnh được người quản trị cập nhật và quản lý nhân công. Trong trường hợp tô-pô mạng thay đổi, người quản trị phải cập nhật lại bảng định tuyến một cách thủ công. Định tuyến tĩnh có một số ứng dụng hữu ích. Định tuyến động có khuynh hướng truyền đạt tất cả các thông tin về một liên mạng. Tuy nhiên, trong trường hợp vì lý do an toàn, chúng ta có thể muốn che dấu một số phần của liên mạng. Định tuyến tĩnh cho phép chúng ta chỉ rõ thông tin muốn tiết lộ. Trong trường hợp chỉ có một

đường đi duy nhất tới mạng, thì chỉ một tuyến tĩnh tới mạng là đủ. Loại mạng này được gọi là mạng cụt (stub network). Cấu hình định tuyến tĩnh cho một mạng cụt tránh được lưu lượng cập nhật định tuyến động.

Định tuyến động hoạt động khác với định tuyến tĩnh. Sau khi người quản trị nhập các lệnh cấu hình để khởi tạo định tuyến động, thông tin về tuyến sẽ được cập nhật tự động mỗi khi nhận được một thông tin mới từ liên mạng. Các thay đổi về tô-pô mạng được trao đổi giữa các bộ định tuyến.



Hình 2.20: Tự động thay đổi tuyến khác khi tuyến đang hoạt động bị lỗi

So với định tuyến tĩnh, định tuyến động hoạt động linh hoạt hơn. Theo bảng định tuyến của bộ định tuyến A, gói có thể tới đích của nó qua bộ định tuyến D. Tuy nhiên, còn có một đường đi sẵn có khác tới đích, đó là đi qua bộ định tuyến B. Khi bộ định tuyến A nhận ra rằng liên kết tới bộ định tuyến D bị lỗi, nó điều chỉnh bảng định tuyến và đường đi tới mạng đích sẽ qua bộ

định tuyến B. Khi liên kết giữa bộ định tuyến A và D được khôi phục, bộ định tuyến A có thể một lần nữa thay đổi bảng định tuyến để chuyển đường đi tới đích là qua bộ định tuyến D.

Các giao thức định tuyến động cũng có thể chuyển lưu lượng từ cùng một phiên làm việc qua nhiều đường đi khác nhau trong mạng để có hiệu suất cao hơn. Tính chất này được gọi là chia sẻ tải (load sharing).

2.3.2. Các giao thức định tuyến

Một liên mạng là sự kết hợp của nhiều mạng được nối với nhau bởi các bộ định tuyến. Khi một gói di chuyển từ nguồn đến đích, nó có thể sẽ đi qua nhiều bộ định tuyến trước khi đến được bộ định tuyến nối trực tiếp với mạng đích.

Bộ định tuyến nhận một gói dữ liệu từ một mạng và chuyển nó tới một mạng khác. Một bộ định tuyến thường được nối tới nhiều mạng. Khi nó nhận được một gói dữ liệu, nó phải chuyển gói đến mạng nào? Quyết định dựa trên sự tối ưu: đường đi sẵn có nào là đường đi tối ưu?

Một *metric* là một giá được gán để chuyển qua một mạng. *Metric* tổng cộng của một tuyến cụ thể bằng tổng *metric* của các mạng trên đường đi. Bộ định tuyến sẽ chọn tuyến có *metric* ngắn nhất (nhỏ nhất).

Metric gắn cho mỗi mạng phụ thuộc vào loại giao thức. Một số giao thức đơn giản, chẳng hạn giao thức thông tin định tuyến (RIP - Routing Information Protocol), coi các mạng ngang nhau. Giá để qua mỗi mạng là như nhau; đó chính là số bước nhảy

(hop count). Do vậy nếu một gói dữ liệu qua 10 mạng để tới đích thì giá tổng cộng là 10.

Các giao thức khác, chẳng hạn giao thức mở đường đi ngắn nhất trước (OSPF - Open Shortest Path First), cho phép người quản trị gán giá để qua một mạng dựa trên các loại dịch vụ yêu cầu. Một tuyến qua một mạng có thể có nhiều giá (metric). Ví dụ, nếu thông lượng tối đa là loại dịch vụ mong muốn thì liên kết vệ tinh có *metric* thấp hơn so với đường cáp quang. Trái lại, nếu độ trễ tối thiểu là loại dịch vụ mong muốn thì đường cáp quang có *metric* thấp hơn so với kết nối vệ tinh. OSPF cho phép mỗi bộ định tuyến có nhiều bảng định tuyến dựa trên các loại dịch vụ yêu cầu.

Một số giao thức khác định nghĩa các *metric* hoàn toàn khác. Trong giao thức cổng biên (BGP - Border Gateway Protocol), tiêu chuẩn chính là chính sách và được người quản trị gán. Chính sách định nghĩa những đường đi nào cần được chọn.

Cho dù sử dụng *metric* nào đi nữa thì bộ định tuyến cũng cần lưu trữ một bảng định tuyến để tham khảo khi chuẩn bị chuyển tiếp một gói dữ liệu. Bảng định tuyến cần chỉ rõ đường đi tối ưu cho gói dữ liệu. Ngày này, liên mạng cần đến những bảng định tuyến động. Bảng định tuyến cần được cập nhật ngay khi có một thay đổi trên liên mạng. Ví dụ, chúng cần được cập nhật khi một tuyến không hoạt động, hoặc khi có một tuyến tốt hơn được tạo ra.

Các giao thức định tuyến được tạo ra để đáp ứng nhu cầu cho các bảng định tuyến động. Một giao thức định tuyến là sự

kết hợp của các luật và thủ tục cho phép các bộ định tuyến trên liên mạng thông báo với nhau các thay đổi. Nó cho phép các bộ định tuyến chia sẻ tất cả những gì chúng biết về liên mạng hoặc về các bộ định tuyến lân cận của chúng. Các giao thức định tuyến cũng bao gồm các thủ tục để kết hợp các thông tin nhận được từ các bộ định tuyến khác.

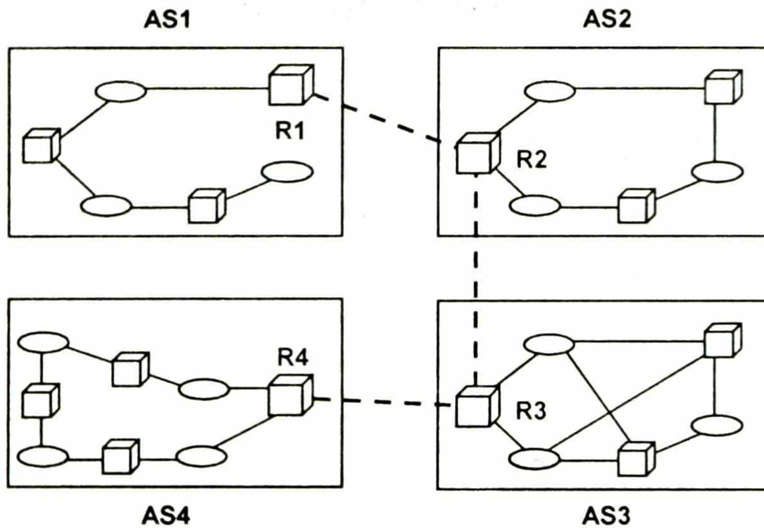
** Định tuyến trong và định tuyến ngoài*

Ngày nay, một liên mạng có thể lớn đến mức một giao thức định tuyến không thể xử lý công việc cập nhật các bảng định tuyến của tất cả các bộ định tuyến. Vì lý do này, liên mạng được chia thành nhiều hệ thống tự trị (AS - Autonomous System). Hệ thống tự trị là một nhóm các mạng và bộ định tuyến chịu một quyền lực quản trị chung. Nó đôi khi còn được gọi là vùng định tuyến (routing domain). Định tuyến bên trong một hệ thống tự trị được gọi là định tuyến trong. Định tuyến giữa các hệ thống tự trị được gọi là định tuyến ngoài. Mỗi hệ thống tự trị có thể chọn một giao thức định tuyến trong để thực hiện định tuyến bên trong hệ thống. Tuy nhiên, thường chỉ có một giao thức định tuyến ngoài được chọn để thực hiện định tuyến giữa các hệ thống tự trị.

Hiện nay có nhiều giao thức định tuyến trong và ngoài đang được sử dụng. Tuy nhiên, các giao thức tiêu biểu và phổ biến nhất đó là hai giao thức định tuyến trong (RIP và OSPF) và một giao thức định tuyến ngoài (BGP).

RIP và OSPF có thể được sử dụng để cập nhật các bảng định tuyến bên trong một hệ thống tự trị. BGP có thể được sử

dụng để cập nhật các bảng định tuyến cho các bộ định tuyến nối các hệ thống tự trị với nhau.



Hình 2.21: Các hệ thống tự trị

Trong hình 2.21, bộ định tuyến R1, R2, R3 và R4 sử dụng một giao thức định tuyến miền ngoài và một giao thức định tuyến miền trong. Các bộ định tuyến khác chỉ sử dụng giao thức định tuyến miền trong. Đường kẻ liền chỉ ra đường truyền thông giữa các bộ định tuyến sử dụng các giao thức định tuyến miền trong. Đường kẻ đứt chỉ ra đường truyền thông giữa các bộ định tuyến sử dụng giao thức định tuyến miền ngoài.

Chương III

CÔNG NGHỆ VOIP

3.1 ĐẶC ĐIỂM VÀ ỨNG DỤNG CỦA VOIP

Ngày nay, các bạn đều ít nhiều hiểu về nguyên lý mạng PSTN và mạng cơ quan. Bạn có thể tự hỏi *"Tại sao phải truyền thoại qua mạng IP?", "Nhân tố nào khiến IP có thể truyền được thông tin thoại hay video?"*.

Nguyên lý cơ bản của VoIP là đóng gói tín hiệu thoại vào trong các gói IP rồi truyền đi như một dạng của thông tin dữ liệu. Vấn đề ở đây là độ trễ, thoại cũng như video là loại thông tin thời gian thực rất nhạy cảm với trễ trong khi IP là một dạng chuyển mạch gói. Chồng giao thức TCP/IP chỉ có khả năng hạn chế được phần nào độ trễ truyền dữ liệu thông qua trường ToS trong tiêu đề IP. Nhưng đó cũng chỉ là "cố gắng tối đa" mà thôi. Như vậy cần phải sử dụng một giao thức lớp trên có khả năng giải quyết vấn đề này. VoIP sử dụng giao thức truyền tải thời gian thực để truyền tải thông tin thời gian thực (RTP/RTCP) như thoại hoặc video mà ta sẽ đề cập ở phần sau.

VoIP sử dụng cơ chế số hóa tín hiệu thoại tốc độ thấp (5,3 kbit/s hoặc 6,3 kbit/s hoặc 8 kbit/s) cho nên nó có thể giảm băng thông yêu cầu đi đáng kể. Ngoài ra, các nhà cung cấp còn đưa ra nhiều giải pháp khác như chủ động bỏ gói tin để đảm bảo

độ trễ cho phép. VoIP áp dụng cơ chế báo hiệu riêng rất "thông minh" cho nên có thể phát triển dễ dàng nhiều dịch vụ tiên tiến kể cả hội nghị thoại truyền hình đa phương tiện.

3.1.1 Ưu điểm của VoIP

Ưu điểm chính của dịch vụ VoIP đối với khách hàng là giá cước rất rẻ so với thoại truyền thống do các cuộc gọi VoIP sử dụng lượng băng thông nhỏ. Trong khi thoại thông thường sử dụng kỹ thuật số hóa PCM theo chuẩn G.711 với lượng băng thông cố định cho một kênh thoại là 64 kbit/s thì VoIP sử dụng kiểu mã hóa tham số nguồn CS-CELP theo chuẩn G.729 (8 kbit/s), G.723 (5,3 hoặc 6,3 kbit/s). Như vậy rõ ràng là lượng băng thông sử dụng giảm một cách đáng kể.

Hơn nữa, thoại truyền thống sử dụng chuyển mạch kênh cho nên lượng băng thông 64 kbit/s cho hướng đi và 64 kbit/s cho hướng về bị chiếm cố định và liên tục trong suốt thời gian diễn ra cuộc nối bất kể hai bên có thực sự đàm thoại với nhau hay không. Trong thực tế khi hai người nói chuyện với nhau thì thường là một người nói và người kia nghe chứ không phải hai bên cùng nói. Vả lại ngay cả đối với người đang nói thì người này cũng có lúc dừng do hết câu hoặc lấy hơi v.v... khi ấy không có thông tin thoại thực sự cần phải truyền đi và người ta gọi là khoảng lặng. Người ta ước tính rằng có ít nhất 70% khoảng thời gian cuộc gọi được gọi là khoảng lặng. VoIP sử dụng cơ chế triệt khoảng lặng cho nên có thể tiết kiệm thêm 70% lượng băng thông "khoảng lặng" này để truyền các dạng thông tin khác.

Nếu bạn để ý chi phí cuộc gọi theo từng phút bạn sẽ thấy lượng tiền tiết kiệm được quả là không nhỏ. Tuy nhiên việc tiết

kiệm này còn tùy thuộc vào vùng địa lý và khoảng cách. Đối với các cuộc gọi nội hạt thì việc tiết kiệm này có vẻ không quan trọng nhưng đối với các cuộc gọi đường dài và nhất là cuộc gọi quốc tế thì nó thật sự rất đáng kể. Điều này được thể hiện ở giá cước mà các nhà cung cấp dịch vụ đưa ra, thông thường giảm còn bằng một phần so với các cuộc gọi quốc tế.

Ưu điểm nữa của VoIP là khả năng dễ dàng kết hợp các loại dịch vụ thoại, dữ liệu và video và như vậy càng tiết kiệm được băng thông do sử dụng chung tài nguyên mạng. Mạng IP đang phát triển một cách mạnh mẽ trên toàn thế giới và càng ngày càng có nhiều ứng dụng "tuyệt vời" đã và đang được phát triển trên nền IP làm cho Internet ngày càng trở nên gần gũi với cuộc sống con người. Nhiều công việc hàng ngày đã dần chuyển sang mạng Internet như mua hàng qua mạng, học tập hay trao đổi thông tin v.v... Giao thức truyền tải thời gian thực RTP/RTCP là công cụ tuyệt vời cho việc truyền tải thoại và video trên mạng IP. Giao thức H.323, SIP và MEGACO được phát triển và đã được chuẩn hóa quốc tế sử dụng cho việc điều khiển cung cấp dịch vụ thông tin đa phương tiện trên nền IP.

3.1.2 Các ứng dụng của VoIP

** Ứng dụng của VoIP cho mạng các mạng công ty*

Đối với khách hàng là các mạng công ty, việc tổ hợp thoại và dữ liệu trong một mạng giúp họ giảm số lượng kênh thuê từ mạng PSTN. Hơn nữa, cơ sở hạ tầng mạng IP (sử dụng các IP phone) ít đòi hỏi việc thay đổi, bổ sung hay di dời hơn trong mạng thoại hay dữ liệu truyền thống. Đó là vì với cùng một cơ

sở hạ tầng, bạn có thể sử dụng các tính năng của mạng dữ liệu như giao thức cấu hình trạm động DHCP (Dynamic Host Configuration Protocol).

DHCP cho phép một thiết bị (PC hay thoại IP) nhận được địa chỉ IP động (tức là địa chỉ IP không cần phải được ấn định tĩnh cho thiết bị). Như vậy, nếu bạn có một thoại IP được ấn định với DHCP, bạn có thể di chuyển máy điện thoại của bạn bất kỳ chỗ nào mà vẫn giữ nguyên số máy. Điều này tương tự như việc máy tính xách tay của bạn từ phòng nọ đến phòng kia mà vẫn có thể truy cập vào cùng máy chủ mạng. Đối với mạng hiện tại việc di chuyển một máy điện thoại khá phức tạp và tốn công sức, thậm chí phải can thiệp đến cấu hình của tổng đài. Nhưng điều này lại rất đơn giản trong mạng VoIP bởi vì khi thoại IP được thiết lập, mạng IP không cần quan tâm bạn đang ở đâu.

Một ưu điểm nữa của VoIP là khả năng cung cấp dịch vụ thư thoại đơn giản với chi phí thấp.

** Trung tâm cuộc gọi thoại gói*

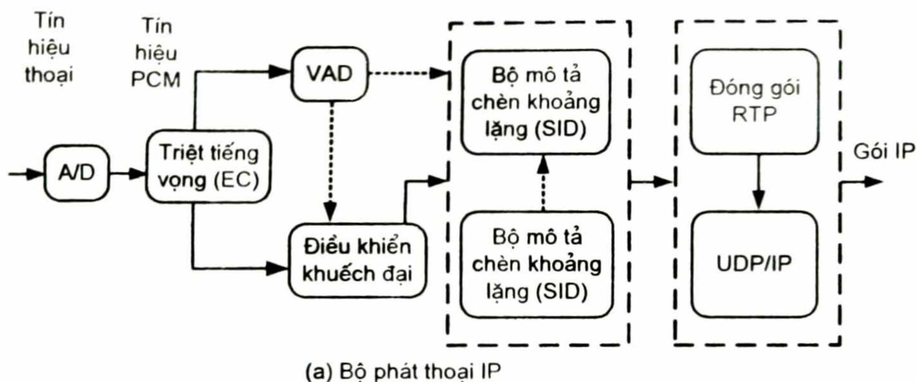
Hầu hết các trung tâm cuộc gọi ngày nay chi phí lớn nhất là xây dựng gắn kết các toà nhà với nhau. Bạn có thể giảm một cách đáng kể chi phí hiện tại cho việc thuê phòng ốc, mua sắm thiết bị đặc chủng (thiết bị định tuyến cuộc gọi, PC v.v...) cần thiết bằng cách sử dụng Trung tâm cuộc gọi thoại gói PTCC (Packet Telephony Call Center). Trong mạng thoại gói, bạn có thể có một nhóm các đại lý (agent) ảo phân bố bất cứ nơi nào bạn muốn mà bạn vẫn có thể thực hiện được những công việc tương tự như các trung tâm cuộc gọi thông thường.

3.2. HỆ THỐNG THU PHÁT VoIP

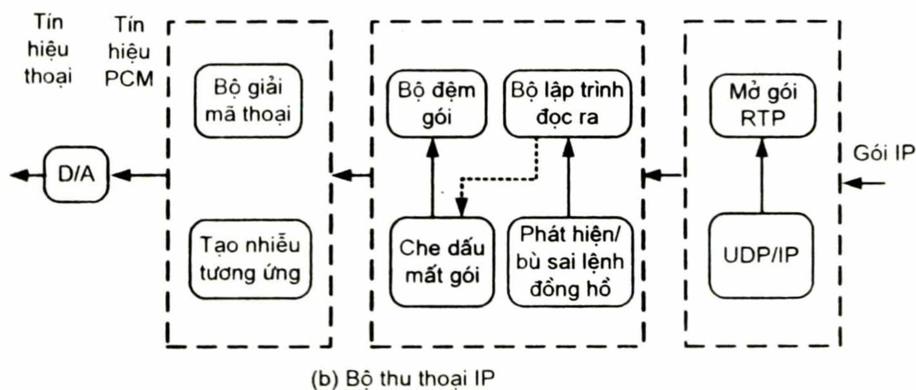
Truyền thông tin thoại thời gian thực từ đầu phát đến đầu thu trong hệ thống thoại IP bao gồm nhiều bước. Ở đầu phát, tín hiệu thoại tương tự được số hóa bằng cách lấy mẫu theo chu kỳ và mã hóa. Tín hiệu số sau đó được xử lý ở bộ triệt tiếng vọng (EC). Bộ phát hiện khoảng lặng được sử dụng để nhận biết chu kỳ khoảng lặng. Trong suốt chu kỳ khoảng lặng, bộ phát không truyền gói hoặc ít ra nó cũng giảm tốc độ bit. Tín hiệu được đóng gói sử dụng bộ mã hóa thoại (coder). Bộ mã hóa thoại cũng có thể nén tín hiệu. Khung thoại sau khi đã đóng gói được truyền qua mạng. Gói giao thức truyền thời gian thực (RTP) được tạo ra bằng cách cộng 12 byte tiêu đề (header) vào khung thoại đã nén. Gói RTP sau đó được đóng gói thành gói UDP ở lớp giao vận và đóng thành gói IP ở lớp mạng. Sau đó, lớp IP sẽ gửi gói vào mạng IP (ví dụ Internet), trong mạng gói được định tuyến và truyền đến đích.

Vì các gói có thể bị mất hoặc trễ trên mạng, nên đầu thu bộ đệm được sử dụng để giảm biến động trễ mạng và lưu các gói cho đến khi chúng được đọc ra. Tại bộ thu, các gói thoại sẽ được mở và giải nén để chuyển đổi chúng thành tín hiệu tương tự ở đầu ra như ban đầu. Các gói không đến được đích hoặc đến quá muộn không được đọc ra thì coi như bị mất. Thuật toán che dấu mất có thể được sử dụng để bù lấp vào các gói đã mất.

Một hệ thống thu phát thoại IP điển hình được minh họa trong hình 3.1.



(a) Bộ phát thoại IP



(b) Bộ thu thoại IP

A/D: Bộ biến đổi tương tự → số

D/A: Bộ biến đổi số → tương tự

VAD: Bộ phát hiện khoảng lặng

Hình 3.1: Hệ thống thu phát thoại IP

3.2.1 Mã hóa thoại

Mã hóa thoại được sử dụng trong truyền thông để đạt được truyền hiệu quả tín hiệu thoại từ đầu thu đến đầu phát. Dải tần tiếng nói có thể nghe được trên điện thoại là 0,1 kHz đến 10 kHz nhưng công suất được phát ra ở dải tần trên 4 kHz là rất nhỏ. 80% toàn bộ năng lượng tiếng nói nằm trong dải từ 300 - 3400 Hz.

Các bộ mã hóa thoại có thể chia làm ba loại là: mã hóa dạng sóng (waveform), mã hóa tham số nguồn (source) và mã hóa lai ghép (hybrid) là kết hợp của hai loại trên.

Nguyên lý của mã hóa dạng sóng là mã hóa dạng sóng của tín hiệu mà đại diện của nó là PCM và ADPCM. Ưu điểm của bộ mã hóa loại này là không phức tạp, giá thành rẻ, độ trễ và công suất tiêu thụ thấp. Tuy nhiên, nó không đảm bảo chất lượng tại tốc độ bit thấp (dưới 16 kbit/s). Bộ mã hóa nguồn khắc phục được nhược điểm này.

Nguyên lý của mã hóa nguồn là mã hóa kiểu phát âm (vocoder), ví dụ như bộ mã hóa dự báo tuyến tính (LPC). Các bộ mã hóa này có thể mã hóa tín hiệu thoại ở tốc độ 2 kbit/s. Đối với mã hóa kiểu phát âm LPC, nó giả thiết rằng: tín hiệu tiếng nói bao gồm cả âm hữu thanh và âm vô thanh. Đối với âm hữu thanh thì nguồn kích thích bộ máy phát âm sẽ là một dãy các xung, còn đối với âm vô thanh thì nó sẽ là một nguồn nhiễu ngẫu nhiên. Trong thực tế có rất nhiều cách để kích thích việc phát âm.

Mã hóa lai ghép (hybrid coder) kết hợp giữa mã hóa dạng sóng và mã hóa nguồn tiêu biểu như dự đoán tuyến tính kích thích mã (CELP - Code Excited Linear Prediction), dự đoán tuyến tính kích thích mã đại số (ACELP - Algebraic Code Excited Linear Prediction), lượng tử hóa gần đúng cực đại đa xung (MP-MLQ).

ITU-T đã chuẩn hóa các kỹ thuật mã hóa tín hiệu thoại như: CELP, MP-MLQ, PCM và ADPCM trong các khuyến nghị G.7xx. Các chuẩn mã hóa thoại thông dụng bao gồm:

- G.711: Kỹ thuật PCM thông thường trong mạng thoại truyền thống PSTN sử dụng công nghệ TDM. Kênh cơ sở tốc độ 64 kbit/s.

- G.726: Mô tả kỹ thuật ADPCM mã hóa với tốc độ 40, 32, 24 và 16 kbit/s.

- G.728: Mô tả kỹ thuật nén thoại CELP ở tốc độ 16 kbit/s với sự thay đổi độ trễ thấp.

- G.729: Mô tả kỹ thuật nén thoại CELP được mã hóa ở tốc độ 8 kbit/s. Chuẩn này có hai dạng là G.729 và G.729 phụ lục A và chúng rất khác nhau về độ phức tạp. Nói chung cả hai đều có thể cung chất lượng dịch vụ thoại tương đương như ở tốc độ 32 kbit/s sử dụng kỹ thuật ADPCM.

- G.723.1: Mô tả kỹ thuật nén thoại hay các tín hiệu âm thanh khác ở tốc độ rất thấp là 5,3 và 6,3 kbit/s. Tốc độ bit cao hơn dựa trên công nghệ MP-MLQ và đương nhiên chất lượng sẽ tốt hơn. Dịch vụ tốc độ thấp dựa trên CELP chất lượng cũng tốt song thiết bị sẽ phức tạp hơn.

Phương pháp để đánh giá chất lượng thoại là Điểm đánh giá trung bình MOS (Mean Opinion Score). Đây là thang điểm 5 để đánh giá mức chất lượng. Bảng 3.1 là một số tiêu chuẩn mã hóa và tốc độ bit cũng như chất lượng.

Bảng 3.1: Các thông số của các bộ mã hóa thoại phổ biến

Chuẩn	Phương pháp mã hóa	Tốc độ bit (kbit/s)	Trễ (ms)	Chất lượng (MOS)
G.711	PCM	64	0,125	4,1
G.726	ADPCM	32	0,125	3,85
G.728	LD-CELP	15	0,625	3,61
G.729	CS-ACELP	8	10	3,92
G.729A	CS-CELP	8	10	3,7
G.723.1	MP-MLQ	6,3	30	3,9
G.723.1	ACELP	5,3	30	3,65

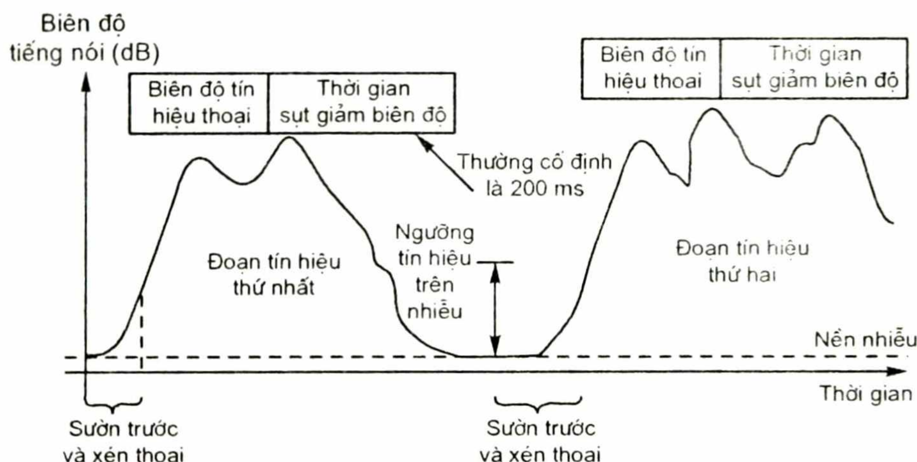
Chi tiết về mã hóa trong VoIP sẽ được trình bày trong mục 3.3.

3.2.2 Phát hiện khoảng lặng (VAD)

Trong đàm thoại, thông thường một người nói và người kia nghe. Như vậy thoại trong mạng PSTN truyền thống sử dụng công nghệ TDM, một cuộc thoại chiếm một kênh hai chiều (gồm một kênh đến và một kênh đi) liên tục trong suốt cuộc nối mà không cần biết ai đang nói và ai đang nghe, thậm chí khi cả hai bên im lặng (khoảng thời gian không nói gọi là khoảng lặng). Nghĩa là ít nhất 50% băng thông bị bỏ phí. Lượng băng thông lãng phí lớn hơn nhiều khi ta tính đến cả khoảng lặng mà người nói ngắt câu hay dừng lấy hơi cũng như dừng để nghĩ. VAD ở bộ phát được sử dụng để tách lời thoại và khoảng lặng. Chỉ có lời thoại được mã hóa và truyền đi để tiết kiệm băng thông.

Hoạt động cơ bản của VAD là so sánh năng lượng tín hiệu trung bình với ngưỡng nhiễu của mỗi khung. Thoại được phát

hiện nếu năng lượng tín hiệu lớn hơn ngưỡng nhiễu. Khi VAD phát hiện năng lượng tín hiệu giảm dưới ngưỡng thì nó nhận biết đó là khoảng lặng và cắt bỏ các xung mẫu thoại.



Hình 3.2: Phát hiện khoảng lặng của tín hiệu thoại (VAD)

Khi VAD phát hiện sự sụt giảm biên độ tín hiệu thoại, nó sẽ đợi một khoảng thời gian nhất định trước khi dừng việc chèn các khung thoại vào trong gói. Khoảng thời gian này được gọi là thời gian sụt giảm biên độ (hangover) và thường là 200 ms.

Một vấn đề nan giải đối với VAD là phát hiện khi thoại bắt đầu (hay kết thúc khoảng lặng), thường thì khi đó một đoạn tín hiệu có thể bị xén mất khi chuyển từ trạng thái từ im lặng sang thoại. Hiện tượng này được xem như bị cắt xén phần đầu, nhìn chung nếu đoạn đầu bị cắt mất lớn hơn 50 ms là người nghe có thể cảm nhận thấy và ảnh hưởng đến chất lượng thoại. Một vấn đề nữa là việc phân biệt tín hiệu thoại và tạp âm nền, khó khăn sẽ xảy ra khi bạn ở trong phòng nhiễu tạp âm và VAD khó có thể phân biệt được tín hiệu thoại với nhiễu nền.

3.2.3 Cơ chế triệt khoảng lặng hoặc truyền gián đoạn

Một ưu điểm chính của VoIP đó là sử dụng băng thông hiệu quả. Gói thoại chỉ cần gửi khi người đàm thoại nói. Trong suốt chu kỳ khoảng lặng khi người đàm thoại nghe, tốc độ bit giảm được đáng kể. Kỹ thuật truyền hai tốc độ (tốc độ khi đàm thoại khác với tốc độ khi nghe) gồm hai cơ chế là triệt khoảng lặng và truyền gián đoạn (khi không đàm thoại chỉ truyền đi tham số của nhiễu nền).

VAD được sử dụng ở bộ phát để phát hiện các đoạn khoảng lặng. Cơ chế triệt khoảng lặng ngừng truyền các gói thoại trong thời gian khoảng lặng. Trong chế độ truyền gián đoạn, các gói mô tả chèn khoảng lặng (SID) bắt đầu được gửi đi và không liên tục qua chu kỳ khoảng lặng. Các khung mô tả chèn khoảng lặng (mô tả để bộ thu căn cứ vào đó để tái tạo lại khoảng lặng) nhỏ hơn khung dữ liệu thoại mã hóa, nó chứa các tham số được sử dụng để tạo nhiễu nền.

Truyền gián đoạn được ưu thích hơn triệt khoảng lặng vì các tham số của nhiễu nền được truyền đến bộ thu, cũng như cơ chế truyền gián đoạn cho phép bộ phát và bộ thu bảo đảm được sự đồng bộ.

3.2.4 Tạo nhiễu nền tương ứng

Trong cơ chế triệt khoảng lặng hoặc truyền gián đoạn, trong chu kỳ khoảng lặng bộ thu không có gói để đọc ra. Nếu bộ thu không có cơ chế gì thì người nghe cảm giác như mất kết nối. Để giải quyết vấn đề này, tại bộ thu phải thực hiện tạo nhiễu nền tương ứng, nhờ đó mà nhiễu nền được tạo và đọc ra trong suốt

thời gian khoảng lặng và người nghe vẫn có cảm giác thực đang đàm thoại. Bộ thu tạo ra nhiễu nền dựa trên các tham số chứa trong các gói mô tả chèn khoảng lặng (SID) nhận được.

3.2.5 Triệt tiếng vọng

Tiếng vọng (echo) trên một cuộc đàm thoại là hiện tượng có thể gây phiền phức từ mức độ nhỏ cho đến không thể chấp nhận được, làm cho người đàm thoại không thể nghe hiểu được nhau. Hiện tượng tiếng vọng gây nên bởi trễ, nó là sự phản xạ của tín hiệu qua mạng và với trễ đủ lớn mà người nghe có thể cảm nhận được. Một số nghiên cứu cho thấy, với trễ vượt quá 32 ms có thể gây khó chịu cho người nghe. Do đó, trong mạng IP người ta phải thực hiện giảm hoặc triệt tiếng vọng (trong trường hợp trễ không thể giảm xuống được nữa). Bộ triệt tiếng vọng ước lượng và loại trừ tiếng vọng từ tín hiệu thu được.

3.2.6 Chương trình đọc ra

Bộ phát thoại IP tạo ra các gói theo các khoảng thời gian đều nhau và gửi chúng qua mạng đến đầu thu. Tính chất của mạng gói kiểu không kết nối là cố gắng tối đa để gửi gói đến đích. Tuy nhiên, nhiều lý do dẫn đến các gói đến đích không còn theo khoảng thời gian đều nhau nữa, thậm trí không theo thứ tự và một số gói còn không bao giờ đến đích được. Sự biến động của khoảng thời gian giãn cách giữa các gói đến được gọi là biến động trễ (jitter).

Vấn đề này được giải quyết bằng cách sử dụng bộ đệm ở đầu thu. Như vậy, các gói đến chưa được đọc ngay (nhất là các

gói đến sớm hơn thời hạn đọc ra nhiều) mà lưu trong bộ đệm trong một khoảng thời gian và sau đó được đọc ra theo thời gian đã lập trình đọc ra cho gói đó. Các gói mà đến sau thời hạn yêu cầu đọc ra đã lập trình thì bộ thu xem như gói đó bị mất. Như vậy, nếu thời gian đệm dài thì sẽ có nhiều gói đến được đọc ra hơn tức là tỷ lệ mất ít hơn. Tuy nhiên, điều này sẽ làm tăng trễ đầu cuối - đầu cuối. Nên cần phải có sự xem xét để cân bằng giữa trễ bộ đệm và mất gói để đảm bảo mất gói ít nhất nhưng vẫn đảm bảo trễ cho phép. Thuật toán tính toán thời gian đệm và đọc ra thích nghi được sử dụng để giải quyết điều này, bằng cách tính toán và điều chỉnh trễ đọc ra theo điều kiện trạng thái mạng hiện thời (theo biến động trễ đối với từng gói).

3.2.7 Che dấu mất gói

Mất gói xảy ra trong thoại IP khi các gói không đến đích như dự định hoặc chúng đến sau thời hạn được đọc ra. Thuật toán che dấu mất gói (PLC - Packet Loss Concealment) được sử dụng ở bộ thu để bù hay che dấu các gói đến muộn hoặc mất. Thuật toán che dấu đơn giản nhất là sử dụng khoảng lặng hoặc nhiễu nền để thay thế cho các gói đã mất. Sử dụng gói tiếp theo (nếu gói đó đến được đích đúng thời hạn và còn tốt) để nội suy hoặc tái tạo lại gói đã mất cũng là giải pháp mà được quan tâm gần đây. Một số kỹ thuật khác như tái tạo dựa trên mô hình mã hóa hay nội suy tham số cũng là cơ chế thực hiện ở đầu thu. Cơ chế thực hiện ở đầu phát kết hợp với đầu thu cũng có các ưu nhược điểm của nó. Sự lựa chọn kỹ thuật nào cho việc khôi phục mất gói cần phải xem xét kỹ, đôi khi người ta còn kết hợp một số kỹ thuật để thực hiện hiệu quả hơn.

3.3. KỸ THUẬT MÃ HÓA TRONG VoIP

Trong mạng điện thoại thông thường tín hiệu thoại được mã hóa PCM theo luật A hoặc μ với tốc độ 64 kbit/s. Cách mã hóa này cho phép khôi phục một cách tương đối trung thực các âm thanh trong dải tần tiếng nói. Tuy nhiên trong một số ứng dụng đặc biệt (ví dụ như truyền tín hiệu thoại trên Internet), âm thanh được truyền với tốc độ thấp hơn. Từ đó đã xuất hiện một số kỹ thuật mã hóa và nén tín hiệu tiếng nói xuống tốc độ thấp cụ thể như G.723.1, G.729A, G.729B và GSM.

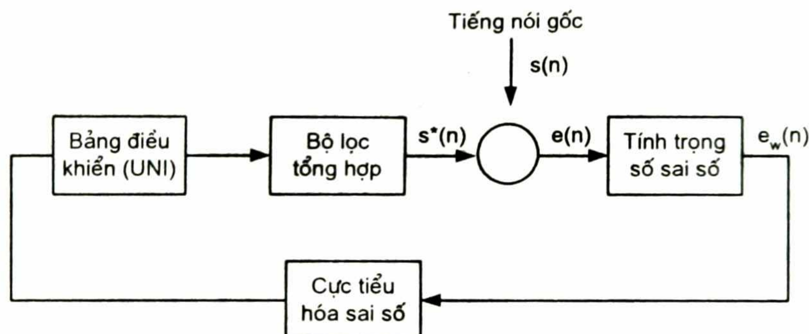
Về cơ bản các bộ mã hóa tiếng nói có ba loại: mã hóa dạng sóng (waveform), mã hóa nguồn (source) và mã hóa lai (hybrid) (nghĩa là kết hợp cả hai loại mã hóa dạng trên). Nguyên lý của mã hóa dạng sóng là mã hóa dạng sóng của tiếng nói. Tại phía phát, bộ mã hóa sẽ nhận các tín hiệu tiếng nói tương tự liên tục và mã thành tín hiệu số trước khi truyền đi. Tại phía thu sẽ làm nhiệm vụ ngược lại để khôi phục tín hiệu tiếng nói. Khi không có lỗi truyền dẫn thì dạng sóng của tiếng nói khôi phục sẽ rất giống với dạng sóng của tiếng nói gốc. Cơ sở của bộ mã hóa dạng sóng là: nếu người nghe nhận được một bản sao dạng sóng của tiếng nói gốc thì chất lượng âm thanh sẽ rất tuyệt vời. Tuy nhiên, trong thực tế, quá trình mã hóa lại sinh ra tạp âm lượng tử (mà thực chất đó là một dạng của méo dạng sóng), song tạp âm lượng tử này thường đủ nhỏ để không ảnh hưởng đến chất lượng tiếng nói thu được. Ưu điểm của bộ mã hóa loại này là: độ phức tạp, giá thành thiết kế, độ trễ và công suất tiêu thụ thấp. Người ta có thể áp dụng chúng để mã hóa các tín hiệu khác như: tín hiệu báo hiệu, số liệu ở dải âm thanh, và đặc biệt với những thiết bị ở điều kiện nhất định thì chúng còn có khả năng mã hóa được cả

tín hiệu âm nhạc. Bộ mã hóa dạng sóng đơn giản nhất là điều chế xung mã (PCM), điều chế Delta (DM)... Tuy nhiên, nhược điểm của bộ mã hóa dạng sóng là không tạo được tiếng nói chất lượng cao tại tốc độ bit dưới 16 kbit/s, bộ mã hóa nguồn khắc phục được nhược điểm này.

Nguyên lý của mã hóa nguồn là mã hóa kiểu phát âm (vocoder), ví dụ như bộ mã hóa dự báo tuyến tính (LPC). Các bộ mã hóa này có thể thực hiện được tại tốc độ bit 2 kbit/s. Hạn chế chủ yếu của mã hóa kiểu phát âm LPC là giả thiết rằng: tín hiệu tiếng nói bao gồm cả âm hữu thanh và âm vô thanh. Do đó, đối với âm hữu thanh thì nguồn kích thích bộ máy phát âm sẽ là một dãy các xung, còn đối với các âm vô thanh thì nó sẽ là một nguồn nhiễu ngẫu nhiên. Trong thực tế có rất nhiều cách để kích thích cơ quan phát âm. Và để đơn giản hóa, người ta giả thiết rằng chỉ có một điểm kích thích trong toàn bộ giai đoạn lên giọng của tiếng nói, dù cho đó là âm hữu thanh.

Vào năm 1982, Atal đã đề xuất một mô hình mới về kích thích, được gọi là kích thích đa xung. Trong mô hình này, không cần biết trước xem đó là âm hữu thanh hay vô thanh, đó có phải là giai đoạn lên giọng hay không. Sự kích thích được mô hình hóa bởi một số xung (thông thường là 4 xung trên 5 ms) có biên độ và vị trí được xác định bằng việc cực tiểu hóa sai lệch, có tính đến trọng số thụ cảm, giữa tiếng nói gốc và tiếng nói tổng hợp. Việc đưa ra mô hình này đã gây chú ý và đó là mô hình đầu tiên của một thế hệ mới của các bộ mã hóa tiếng nói phân tích bằng tổng hợp. Chúng có khả năng cho tiếng nói chất lượng cao tại tốc độ bit khoảng 10 kbit/s và có thể chỉ còn 4,8 kbit/s. Tín hiệu kích thích sẽ được tối ưu hóa một cách kỹ lưỡng và người ta sử dụng kỹ thuật mã hóa dạng sóng để mã hóa tín hiệu kích thích này một

cách có hiệu quả. Hình 3.3 đưa ra mô hình tổng quát của mã hóa tiếng nói theo phương pháp LPC phân tích bằng tổng hợp.



Hình 3.3: Mô hình mã hóa tiếng nói theo phương pháp LPC

3.3.1 Nguyên lý bộ mã hóa CS-ACELP

Tín hiệu PCM 64 kbit/s đầu vào (theo luật A hoặc luật μ) qua bộ mã hóa thuật toán CS-ACELP, được lấy mẫu tại tần số 8 kHz, sau đó qua bộ chuyển đổi thành tín hiệu PCM đều 16 bit đưa tới đầu vào bộ mã hóa. Tín hiệu đầu ra bộ giải mã sẽ được chuyển đổi thành tín hiệu PCM (theo qui luật A hoặc μ) theo đúng tín hiệu đầu vào. Các đặc tính đầu vào/đầu ra khác, giống như của tín hiệu PCM 64 kbit/s (theo Khuyến nghị ITU G.711), sẽ được chuyển đổi thành tín hiệu PCM đều 16 bit tại đầu vào bộ mã hóa, hoặc tín hiệu PCM đều 16 bit sẽ được chuyển đổi thành tín hiệu ra PCM theo đúng qui luật của tín hiệu đầu vào ở bộ giải mã sẽ được trình bày chi tiết sau.

3.1.1.1 Nguyên lý chung của bộ mã hóa

Bộ mã hóa CS-ACELP dựa trên cơ sở của bộ mã dự báo tuyến tính kích thích mã CELP.

Bộ mã hóa CS-ACELP thực hiện trên các khung tiếng nói chu kỳ 10 ms tương đương với 80 mẫu tại tốc độ lấy mẫu là 8000 mẫu/giây. Cứ mỗi một khung 10 ms, tín hiệu tiếng nói lại được phân tích để trích lấy các tham số của bộ mã CELP (đó là: các hệ số của bộ lọc dự báo thích ứng, chỉ số các bảng mã cố định và bảng mã thích ứng cùng với các tăng ích của các bảng mã). Các tham số này sẽ được mã hóa và truyền đi. Sự phân bố bit của các tham số mã hóa được trình bày ở bảng 3.2.

Bảng 3.2: Sự phân bố bit của các tham số của thuật toán CS-ACELP tốc độ 8 kbit/s (khung 10 ms)

Tham số	**Từ mã**	**Số bit trong khung con 1**	**Số bit trong khung con 2**	**Tổng số bit trong 1 khung**
Các cặp vạch phổ	L0,L1,L2,L3			18
Độ trễ mã thích ứng	P1, P2	8	5	13
Độ chẵn lẻ trễ bước	P0	1		1
Chỉ số mã cố định	C1, C2	13	13	26
Dấu mã cố định	S1, S2	4	4	8
Các độ khuếch đại mã (bước 1)	GA1, GA2	3	3	6
Các độ khuếch đại mã (bước 2)	GB1, GB2	4	4	8
Tổng cộng				80

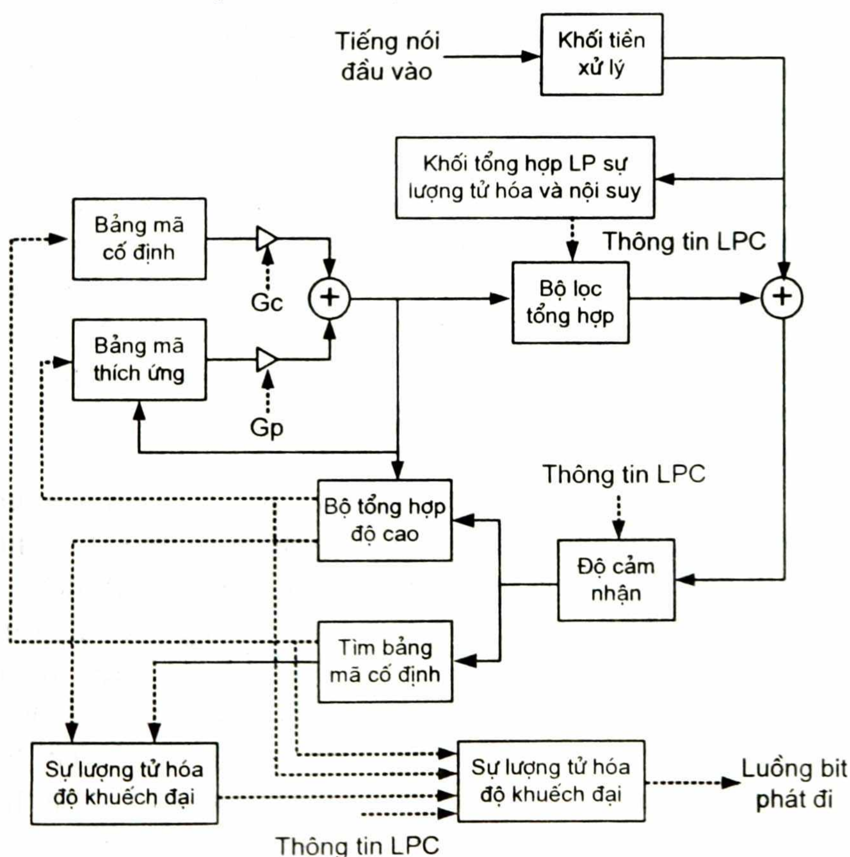
Tại phía thu: sử dụng các tham số này để khôi phục các tham số tín hiệu kích thích và các tham số của bộ lọc tổng hợp. Tín hiệu tiếng nói sẽ được khôi phục bằng cách lọc các tham số tín hiệu kích thích này thông qua bộ lọc tổng hợp ngắn hạn.

Bộ lọc tổng hợp ngắn hạn dựa trên cơ sở bộ lọc dự báo tuyến tính LP bậc 10. Bộ lọc tổng hợp dài hạn, hay bộ lọc tổng

hợp độ cao dùng cho việc làm tròn mã thích ứng. Sau khi khôi phục, nhờ bộ lọc sau tiếng nói sẽ được làm tăng độ trung thực.

3.3.1.2 Nguyên lý bộ mã hóa CS-ACELP

Sơ đồ khối bộ mã hóa được mô tả trên hình 3.4.



Hình 3.4: Sơ đồ nguyên lý của bộ mã hóa CS-ACELP

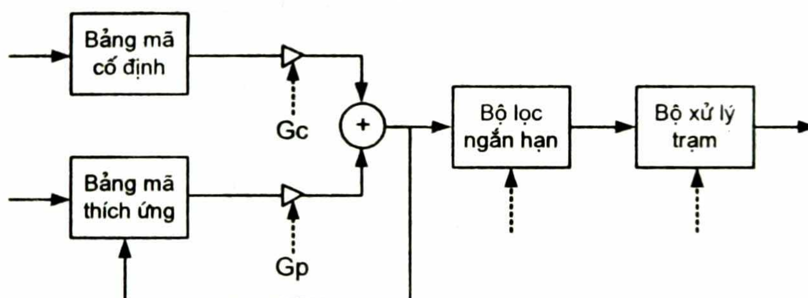
Tín hiệu đầu vào đưa qua bộ tiền xử lý, bộ này có hai chức năng: lọc thông cao và tính toán tín hiệu. Tín hiệu đầu ra bộ tiền xử lý là tín hiệu đầu vào của các khối tổng hợp tiếp sau. Sự tổng hợp dự báo tuyến tính (LP) được thực hiện một lần trong một

khung 10 ms để tính các hệ số của bộ lọc dự báo tuyến tính (LP). Các hệ số này được biến đổi thành các cặp vạch phổ (LSP) và được lượng tử bằng phương pháp lượng tử hóa véc tơ dự báo hai bước (VQ) 18 bit. Tín hiệu kích thích được lựa chọn bằng cách cực tiểu hóa sai số, có tính đến trọng số thụ cảm giữa tiếng nói gốc và tiếng nói tổng hợp. Các tham số kích thích (gồm bảng mã cố định và bảng mã thích ứng) được xác định qua từng khung con 5 ms (tương đương với 40 mẫu). Các hệ số của bộ lọc LP đã được lượng tử và chưa được lượng tử được sử dụng cho phân khung thứ 2, còn tại phân khung thứ nhất các hệ số của bộ lọc LP đã được nội suy sẽ được sử dụng (trong cả hai trường hợp đã lượng tử và chưa lượng tử). Độ trễ bước mạch vòng hở sẽ được tính toán một lần trong một khung 10 ms dựa trên độ lớn tín hiệu thoại. Sau đó các phép tính này sẽ lặp lại trong từng phân khung tiếp theo. Tín hiệu ban đầu x(n) được tính bằng cách lọc độ dư LP thông qua bộ lọc tổng hợp W(z)/A(z). Trạng thái ban đầu của bộ lọc này là tín hiệu lỗi giữa tín hiệu độ dư LP và tín hiệu kích thích. Sự phân tích bước của mạch vòng đóng sẽ được thực hiện sau đó (để tìm độ trễ mã thích ứng và độ khuếch đại), dùng tín hiệu ban đầu x(n) và đặc tuyến xung h(n), bằng cách làm tròn giá trị độ trễ bước của mạch vòng hở. Độ trễ bước được mã hóa bằng mã 8 bit trong phân khung thứ nhất, độ vi sai của độ trễ được mã hóa bằng mã 5 bit trong phân khung thứ 2. Tín hiệu x′(n) là hiệu số của 2 tín hiệu: tín hiệu ban đầu x(n) và tín hiệu mã thích ứng (tín hiệu mã cố định). Tín hiệu này được dùng trong việc tìm tín hiệu kích thích tối ưu. Giá trị kích thích mã cố định được mã hóa bằng mã đại số 17 bit (trong đó: Chỉ số bảng mã cố định được mã hóa bằng từ mã C1, C2 (13 bit); Dấu bảng mã cố định được mã hóa bằng từ mã S1, S2 - 3 bit). Các độ khuếch đại bảng mã cố định và bảng mã thích ứng được lượng tử hóa bằng véc-tơ 7 bit (Trong đó: ở bước 1 được mã hóa bằng từ

mã GA1, GA2 - 3 bit: ở bước 2 được mã hóa bằng từ mã GB1, GB2 - 4 bit). Tại đây sử dụng sự dự đoán trung bình động MA cho độ khuếch đại mã cố định. Cuối cùng, dựa vào các bộ nhớ lọc sẽ xác định được tín hiệu kích thích.

3.3.1.3 Nguyên lý bộ giải mã CS-ACELP

Sơ đồ khối của bộ giải mã được mô tả trong hình 3.5.



Hình 3.5: Sơ đồ nguyên lý của bộ giải mã CS-ACELP

Đầu tiên, các chỉ số của các tham số được trích ra từ luồng bit thu. Các chỉ số này sẽ được giải mã để thu lại các tham số của bộ mã hóa trong 1 khung tiếng nói 10 ms. Các tham số đó là: các hệ số LSP, 2 phần độ trễ bước (độ trễ bước và độ vi sai của độ trễ bước), 2 véc-tơ bảng mã cố định (chỉ số mã cố định và chỉ số bảng mã cố định), 2 tập hợp độ khuếch đại bảng mã cố định và bảng mã thích ứng. Các hệ số LSP được nội suy và được chuyển đổi thành các hệ số bộ lọc LP cho mỗi phân khung. Sau đó, cứ mỗi phân khung thực hiện các bước tiếp theo:

- Giá trị kích thích được khôi phục là tổng của véc-tơ bảng mã cố định và bảng mã thích ứng nhân với các giá trị khuếch đại tương ứng của chúng.

- Tiếng nói được khôi phục bằng cách lọc giá trị kích thích này thông qua bộ lọc tổng hợp LP.

- Tín hiệu tiếng nói khôi phục đưa qua bước xử lý trạm, bao gồm bộ lọc thích ứng dựa trên cơ sở các bộ lọc tổng hợp ngắn hạn và dài hạn, sau đó qua bộ lọc thông cao và bộ nâng tín hiệu.

3.3.2 Chuẩn nén G.729A

G.729A là thuật toán mã hóa tiếng nói tiêu chuẩn cho thoại và số liệu đồng thời số hóa (DSVD). G.729A là sự trao đổi luồng bit với G.729, có nghĩa là: tín hiệu được mã hóa bằng thuật toán G.729A có thể được giải mã thông qua thuật toán G.729 và ngược lại. Giống như G.729, nó sử dụng thuật toán dự báo tuyến tính mã kích thích đại số được cấu trúc liên kết (CS-ACELP) với các khung 10 ms. Tuy nhiên một vài thuật toán thay đổi sẽ được giới thiệu mà kết quả của các thuật toán này làm giảm 50% độ phức tạp.

Nguyên lý chung của bộ mã hóa và giải mã của thuật toán G.729A giống với G.729. Các thủ tục lượng tử hóa và phân tích LP của các độ khuếch đại bảng mã cố định và thích ứng giống như G.729. Các thay đổi thuật toán chính so với G.729 sẽ tổng kết như sau:

- Bộ lọc trọng số thụ cảm sử dụng các tham số bộ lọc LP đã lượng tử và được biểu diễn là $W(z) = A(z)/A(z/\gamma)$ với giá trị cố định $\gamma = 0,75$.

- Phân tích độ lên giọng mạch vòng hở được đơn giản hóa bằng cách sử dụng phương pháp ***decimation*** (có nghĩa là trích 10 lấy 1) trong khi tính sự tương quan của tiếng nói trọng số.

- Các tính toán phản ứng xung của bộ lọc tổng hợp trọng số $W(z)/A(z)$, của tín hiệu ban đầu và việc thiết lập trạng thái ban đầu của bộ lọc được đơn giản hóa bằng cách thay thế $W(z)$ bằng $1/A(z/\gamma)$.

- Việc tìm bảng mã thích ứng được đơn giản hóa. Việc tìm sẽ cực đại hóa giá trị tương quan giữa kích thích trước và tín hiệu ban đầu lọc trước (năng lượng của kích thích trước lọc là không đáng kể).

- Việc tìm bảng mã cố định được đơn giản hóa. Thay vì tìm tập trung ở mạch vòng tổ ong, giải pháp tìm sơ đồ hình cây độ sâu trước được sử dụng.

- Tại bộ giải mã, bộ lọc sau sẽ được đơn giản bằng cách sử dụng chỉ các độ trễ nguyên.

Bảng 3.3: Các thông số WMOPS và MIPS của G.729 và G.729A

Chức năng	WMOPS		C50 MIPS	
	G.729	**G.729A**	**G.729**	**G.729A**
Tiền xử lý	0,20	0,20	0,226	0,226
Phân tích LP	1,63	1,28	1,957	1,696
Lượng tử hóa và nội suy LSP	0,95	0,95	1,390	1,390
Biến đổi LSP thành A(z) và trọng số	0,30	0,12	0,461	0,173
Lên giọng mạch vòng hở	1,45	0,82	1,563	0,955
Lên giọng mạch vòng đóng	2,83	1,55	3,453	1,778
Bảng mã đại số	6,35	1,86	8,406	3,046
Lượng tử hóa các độ khuếch đại	0,46	0,46	0,643	0,643
Tìm kích thích và cập nhật bộ nhớ	0,21	0,08	0,278	0,112
Tổng cộng (mã hóa)	*14,38*	*7,32*	*18,377*	*10,019*
Giải mã	0,68	0,68	1,133	1,133
Bộ lọc sau	2,13	0,73	2,539	1,000
Xử lý sau	0,22	0,22	0,266	0,226
Tổng cộng (giải mã)	*3,03*	*1,63*	*3,938*	*2,399*
Tổng cộng (mã hóa + giải mã)	*17,41*	*8,95*	*22,315*	*12,418*

MIPS: (Million Instructions Per Second): triệu câu lệnh trên 1 giây
WMOPS: (Weighted Million Operations Per Second): triệu thao tác trên 1 giây.

Cả hai bộ mã hóa G.729 và G.729A đã được thử nghiệm trên vi mạch T1 TMS320C50 DSP. Trong thử nghiệm, thuật toán mã hóa song công G.729A yêu cầu 12,4 MIPS, trong khi G.729 yêu cầu 22,3 MIPS. Việc giảm độ phức hợp của cả hai bộ mã hóa G.729 và G.729A được đưa ra trong bảng 3.3 cho cả hai phần mã hóa và giải mã. Độ phức tạp ở đây được thể hiện qua 2 số hạng: C50 MIPS và WMOPS của thuật toán cơ sở. Về yêu cầu bộ nhớ: G.729A yêu cầu ít hơn 2K RAM và 10K ROM trong khi G.729 yêu cầu khoảng 2K RAM và 11K ROM. Hiển nhiên rằng: sử dụng G.729A giảm được khoảng 50% độ phức tạp so với sử dụng G.729, với việc giảm một ít chất lượng trong trường hợp 3 bộ đôi (mã hóa/giải mã) và trong trường hợp có tạp âm nền.

3.3.3 Chuẩn nén G.729B

G.729B đưa ra một nguyên lý nén khoảng lặng tốc độ bit thấp được thiết kế và tối ưu hóa để làm việc chung được với cả G.729 và G.729A phức tạp thấp. Để đạt được việc nén khoảng lặng tốc độ bit thấp chất lượng tốt, một mô-đun bộ dò hoạt động thoại khung cơ bản là yếu tố cần thiết để dò các khung thoại không tích cực, gọi là các khung tạp âm nền hoặc khung im lặng. Đối với các khung thoại không tích cực đã dò được này, một mô-đun truyền gián đoạn đo sự thay đổi theo thời gian của đặc tính tín hiệu thoại không tích cực và quyết định xem có một khung mô tả thông tin im lặng mới có thể được gửi đi để duy trì chất lượng tái tạo của tạp âm nền tại đầu cuối thu được không? Nếu có một khung như thế được yêu cầu, các tham số năng

lượng và phổ mô tả các đặc tính cảm nhận được của tạp âm nền được mã hóa và truyền đi một cách hiệu quả dùng 15 bit/khung. Tại đầu cuối thu, mô-đun tạo tạp âm phù hợp sẽ tạo tạp âm nền đầu ra sử dụng tham số cập nhật đã phát hoặc các tham số đã có trước đó. Tạp âm nền tổng hợp đạt được bằng cách lọc dự báo tuyến tính tín hiệu kích thích tạp giả trắng được tạo ra trong nội bộ của mức điều khiển. Phương pháp mã hóa tạp âm nền tiết kiệm tốc độ bit cho tiếng nói mã hóa tại tốc độ bit trung bình thấp 4 kbit/s trong cuộc đàm thoại tiếng nói bình thường để duy trì chất lượng tái tạo.

Đối với các ứng dụng DSVD (Digital Simultaneous Voice and Data: thoại và số liệu đồng thời số hóa) và độ nhạy tốc độ bit khác, G.729B là một điều kiện tối cần thiết để giảm tốc độ bit hơn nữa bằng cách sử dụng công nghệ nén im lặng. Khi không có tiếng nói, tốc độ bit có thể giảm, giải phóng dung lượng kênh cho các ứng dụng xảy ra đồng thời, ví dụ như các đường truyền tiếng khác trong điện thoại tế bào đa truy nhập phân kênh theo mã/theo thời gian (TDMA/CDMA) hoặc truyền số liệu đồng thời. Một phần đáng kể trong các cuộc đàm thoại thông thường là im lặng, trung bình lên tới 60% của một cuộc đàm thoại 2 chiều. Trong suốt quá trình im lặng, thiết bị đầu vào tiếng, ví dụ như tai nghe, sẽ thu thông tin từ môi trường ồn. Mức và đặc tính ồn có thể thay đổi đáng kể, từ một phòng im lặng tới đường phố ồn ào hoặc từ một chiếc xe ô tô chuyển động nhanh. Tuy nhiên, hầu hết các nguồn tạp âm thường mang ít thông tin hơn thông tin tiếng, vì vậy, trong các chu kỳ không tích cực tỷ số nén sẽ cao hơn. Nhiều ứng dụng điển hình, ví dụ hệ thống toàn cầu đối với

điện thoại di động GSM, sử dụng việc dò tìm chu kỳ im lặng và chèn tạp âm phù hợp để tạo được hiệu quả mã hóa cao hơn.

Xuất phát từ quan niệm về dò tìm im lặng và chèn tạp âm phù hợp dẫn tới các công nghệ mã hóa tiếng mẫu kép. Các mẫu khác nhau bởi tín hiệu đầu vào, được biểu thị là: thoại tích cực đối với tiếng nói và là thoại không tích cực đối với khoảng lặng hoặc tạp âm nền, được xác định bởi sự phân loại tín hiệu. Sự phân loại này có thể được thực hiện bên trong hoặc bên ngoài bộ mã hóa tiếng nói. Bộ mã hóa tiếng toàn tốc có thể có tác dụng trong quá trình tiếng thoại tích cực, nhưng có một nguyên lý khác được dùng đối với tín hiệu thoại không tích cực, sử dụng bit ít hơn và tạo ra tỷ số nén trung bình cao hơn. Sự phân loại này được gọi chung là bộ dò hoạt động thoại (VAD: Voice Activity Detector) và đầu ra của bộ này gọi là mức hoạt động thoại. Mức hoạt động thoại là 1 khi có hoạt động thoại và là 0 khi không có hoạt động thoại.

Thuật toán VAD và bộ mã hóa tiếng nói không tích cực, giống với các bộ mã hóa G.729 và G.729A, được thực hiện trên các khung của tiếng nói đã được số hóa. Để phù hợp, kích thước các khung giống nhau được dùng cho mọi sơ đồ và không có độ trễ thêm vào nào được tạo ra bởi thuật toán VAD hoặc bộ mã hóa thoại không tích cực. Đầu vào bộ mã hóa tiếng nói là tín hiệu tiếng nói đến đã được số hóa. Với mỗi khung tiếng nói đầu vào, VAD đưa ra mức hoạt động thoại, mức này được dùng như một chuyển mạch giữa các bộ mã hóa thoại tích cực và thoại không tích cực. Khi bộ mã hóa thoại tích cực có tác dụng, luồng bit thoại tích cực sẽ gửi tới bộ giải mã tích cực cho mỗi khung.

Tuy nhiên, trong các chu kỳ không tích cực, bộ mã hóa thoại không tích cực có thể được chọn để gửi các thông tin mới nhất gọi là bộ mô tả việc chèn khoảng lặng (SID: Silence Insertion Descriptor) tới bộ giải mã không tích cực hoặc không gửi gì cả. Kỹ thuật này có tên là truyền gián đoạn (DTX: Discontinuous Transmission). Với mỗi khung, đầu ra của mỗi bộ giải mã được dùng làm tín hiệu khôi phục.

3.3.3 Chuẩn nén G.723.1

Khuyến nghị G.723.1 đưa ra một bộ mã hóa tiêu chuẩn dùng để nén tín hiệu tiếng nói hoặc các tín hiệu audio khác của các dịch vụ đa phương tiện tại tốc độ rất thấp, giống với phần tiêu chuẩn của họ H.323.

Về tốc độ bit: bộ mã hóa này có 2 tốc độ bit: 5,3 kbit/s và 6,3 kbit/s. Bộ mã hóa có tốc độ bit cao hơn sẽ có chất lượng cao hơn. Bộ mã hóa có tốc độ bit thấp hơn đem lại chất lượng khá tốt và cộng thêm tính linh hoạt, cung cấp sự mềm dẻo cho các nhà thiết kế hệ thống. Bộ mã hóa và giải mã bắt buộc phải có cả 2 tốc độ bit này. Chúng có thể chuyển mạch được giữa 2 tốc độ bit. Khi tín hiệu là phi thoại thì có thể lựa chọn một tốc độ bit biến thiên để truyền không liên tục và điều khiển những khoảng trống.

Tín hiệu đầu vào có thể của bộ mã hóa này tối ưu hóa tín hiệu tiếng nói với chất lượng cao tại các tốc độ bit đã nói ở trên với một hạn chế nhất định về độ phức tạp. Bộ mã hóa này dùng để mã hóa tiếng nói và các tín hiệu audio khác với các khung dùng kỹ thuật mã hóa phân tích bằng tổng hợp dự báo tuyến tính. Tín hiệu kích thích, đối với bộ mã hóa tốc độ bit cao hơn, là

lượng tử hóa đúng cực đại đa xung (MP-MLQ) và đối với bộ mã hóa có tốc độ bit thấp hơn, là dự đoán tuyến tính kích thích mã đại số (ACELP). Kích thích khung là 30 ms, cộng thêm 7,5 ms look-ahead, tạo ra trễ xử lý thuật toán tổng cộng là 37,5 ms. Toàn bộ trễ thêm vào bộ mã hóa là tổng của: Trễ xử lý, trễ truyền dẫn trên các đường truyền thông tin và trễ đệm của các giao thức ghép kênh.

3.3.3.1 Nguyên lý bộ mã hóa G.723.1

Tín hiệu PCM 64 kbit/s đầu vào (theo qui luật A hoặc μ) qua bộ mã hóa này, được lấy mẫu tại tần số 8 kHz, sau đó qua bộ chuyển đổi thành tín hiệu PCM đều 16 bit đưa tới đầu vào bộ mã hóa. Tín hiệu đầu ra bộ giải mã sẽ được chuyển đổi thành tín hiệu PCM (theo qui luật A hoặc μ) theo đúng tín hiệu đầu vào. Các đặc tính đầu vào/đầu ra khác, giống như của tín hiệu PCM 64 kbit/s (theo khuyến nghị ITU G.711), sẽ được chuyển đổi thành tín hiệu PCM đều 16 bit tại đầu vào bộ mã hóa, hoặc tín hiệu PCM đều 16 bit sẽ được chuyển đổi thành tín hiệu ra PCM theo đúng qui luật của tín hiệu đầu vào ở bộ giải mã sẽ được trình bày chi tiết sau.

Bộ mã hóa dựa trên nguyên lý bộ mã hóa phân tích bằng tổng hợp dự báo tuyến tính và cố gắng cực tiểu hóa sai số có tính trọng số thụ cảm. Bộ mã hóa thực hiện theo từng khung 240 mẫu. Điều này tương đương với chu kỳ khung là 30 ms và tần số lấy mẫu là 8 kHz. Tại mỗi khối, đầu tiên tín hiệu được đưa qua bộ lọc thông cao để loại bỏ thành phần tín hiệu một chiều DC và sau đó được chia thành 4 khung con. Với mỗi khung con, sử

dụng tín hiệu đầu vào chưa xử lý để tính toán bộ lọc mã hóa dự báo tuyến tính bậc 10 (LPC). Bộ lọc LPC của khung con cuối cùng sẽ được lượng tử hóa bằng phương pháp lượng tử hóa véctơ phân chia dự báo (PSVQ). Các hệ số LPC chưa được lượng tử sẽ được dùng để khôi phục bộ lọc trọng số thụ cảm ngắn hạn.

Với mỗi hai phân khung (120 mẫu), sẽ sử dụng tín hiệu tiếng nói trọng số để tính toán chu kỳ lên giọng tiếng nói mạch vòng kín, L_{OL}. Chu kỳ lên giọng tiếng nói được tính trong khoảng từ 18 đến 142 mẫu. Sau đó tín hiệu tiếng nói sẽ được xử lý theo từng phân khung cơ bản 60 mẫu.

Sử dụng đánh giá chu kỳ lên giọng tiếng nói trước để khôi phục bộ lọc dạng ồn họa ba. Phản ứng xung được tạo bởi việc đấu nối bộ lọc tổng hợp LPC, bộ lọc có tính trọng số thụ cảm formant và bộ lọc dạng tạp âm họa ba. Người ta sử dụng phản ứng xung này cho các phép tính toán tiếp sau.

Bộ dự đoán chu kỳ lên giọng mạch vòng kín được tính toán bằng cách sử dụng đánh giá chu kỳ lên giọng, L_{OL}, và phản ứng xung. Người ta sử dụng bộ dự đoán lên giọng bậc 5. Chu kỳ lên giọng sẽ được tính là gần đúng giá trị vi sai nhỏ của đánh giá lên giọng mạch vòng hở. Thành phần thêm vào bộ dự đoán lên giọng sau đó sẽ được loại bỏ khỏi véc-tơ ban đầu. Cả hai giá trị chu kỳ lên giọng và giá trị vi sai của nó sẽ được truyền về phía bộ giải mã.

Cuối cùng, các thành phần không được dự đoán của tín hiệu kích thích sẽ được lấy gần đúng. Đối với bộ mã hóa có tốc độ bit cao, người ta sử dụng giá trị kích thích lượng tử hóa gần đúng

cực đại đa xung (MP-MLQ), và đối với bộ mã hóa có tốc độ bit thấp, người ta sử dụng giá trị kích thích mã đại số (ACELP).

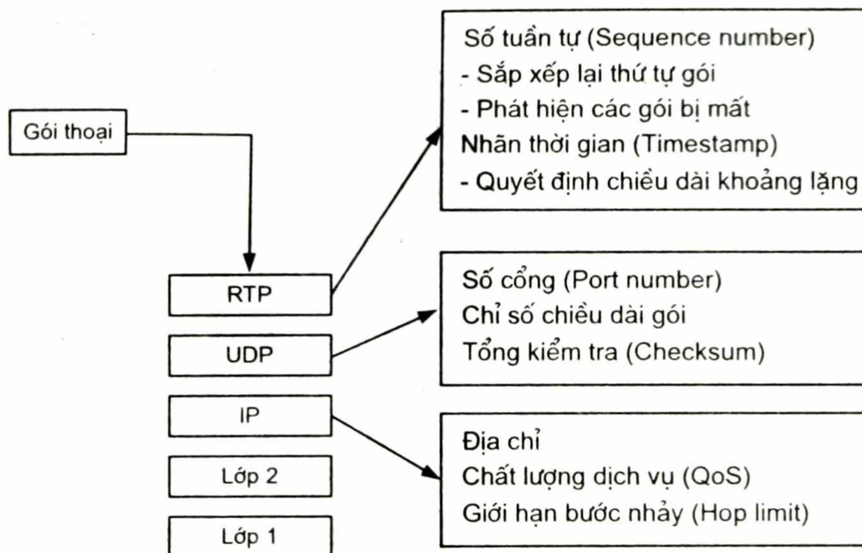
3.3.3.2 Nguyên lý bộ giải mã G.723.1

Bộ giải mã được thực hiện trên nguyên lý cơ bản từng khung. Đầu tiên các chỉ số của bộ lọc LPC sẽ được giải mã, sau đó bộ giải mã sẽ khôi phục bộ lọc tổng hợp LPC. Đối với mỗi phân khung, cả hai giá trị kích thích bảng mã cố định và giá trị kích thích bảng mã thích ứng sẽ được giải mã và đưa tới đầu vào bộ lọc tổng hợp LPC. Bộ lọc sau thích ứng bao gồm formant và bộ lọc sau lên giọng phía sau-phía trước (forward-backward). Tín hiệu kích thích sẽ được đưa tới đầu vào bộ lọc sau lên giọng, đầu ra bộ lọc sau lên giọng được đưa tới đầu vào bộ lọc tổng hợp và đầu ra bộ lọc tổng hợp sẽ được đưa tới đầu vào bộ lọc sau formant.

3.4. GIAO THỨC TRUYỀN TẢI THỜI GIAN THỰC (RTP/RTCP)

Phương tiện truyền thông trong thoại IP được thực hiện với giao thức truyền thời gian thực (RTP). RTP cung cấp truyền dữ liệu thời gian thực từ đầu cuối đến đầu cuối như audio, video. Gói RTP bao gồm một tiêu đề RTP và phần tải tin (payload). Các gói RTP được mang trong gói UDP/IP. RTP không có sự bảo đảm về chất lượng và cũng không có cơ chế dành trước tài nguyên cho kết nối. Tuy nhiên, tiêu đề RTP có trường số tuần tự (sequence number) để đầu thu có thể phát hiện mất gói xảy ra và các gói đến không đúng thứ tự. Hơn nữa, tiêu đề gói RTP có trường timestamp (nhãn thời gian) để tính toán trễ mạng và biến động trễ cũng như đảm bảo việc đọc ra đồng bộ.

Để truyền tải gói thoại chồng giao thức RTP/UDP/IP được sử dụng như hình 3.6 dưới đây.



Hình 3.6: Chồng giao thức RTP/UDP/IP sử dụng cho truyền tải thoại

3.4.1. Giao thức truyền thời gian thực (RTP)

Do đặc tính nhạy cảm với độ trễ của tín hiệu thoại nên UDP/IP được lựa chọn làm giải pháp hợp lý để truyền thông tin thoại. Ngoài ra, IETF chấp nhận sử dụng RTP đối với lưu lượng thời gian thực hoặc lưu lượng nhạy cảm với trễ. Thoại được mã hóa và đóng gói trong gói RTP/UDP/IP. Khi một trạm muốn gửi một gói tin, nó tiến hành việc định dạng luồng thông tin đa phương tiện đó để đóng gói, thêm phần nhãn tiêu đề tương ứng vào mỗi gói (chẳng hạn tiêu đề RTP) rồi đặt chúng vào trường tải tin lớp thấp hơn là UDP và IP. Gói sau đó được gửi trên mạng, có thể theo kiểu đa hướng hay đơn hướng tới trạm đích.

RTP được thiết kế để hỗ trợ một số dịch vụ quan trọng trong việc cung cấp thông tin với các yêu cầu thời gian thực, đặc biệt là lưu lượng thoại và video. RTP cố gắng đưa một số quy tắc và khả năng dự đoán trước cho các ứng dụng mà sử dụng Internet để truyền lưu lượng nhạy cảm về thời gian.

1. Nhãn thời gian. Sự biến đổi của độ trễ giữa người gửi và người nhận trong mạng sử dụng công nghệ gói (chẳng hạn như Internet) luôn đưa đến kết quả là gói đến đích trong trạng thái từng "khối". Điều đó có nghĩa là một số lượng lớn các gói đến cùng một lúc, sau đó không có gói nào rồi lại đến giai đoạn các gói đến hàng loạt và cứ thế tiếp tục. Quá trình đó tạo ra biến động trễ (jitter), nó không giống như môi trường tối ưu cho việc quy định mạng trong đó dữ liệu để có một độ trễ nhất định. RTP giải quyết vấn đề này bằng cách sử dụng trường 32 bit để xác định thời điểm mà trường dữ liệu bắt đầu được lấy mẫu, ngẫu nhiên xác định gói dữ liệu đầu tiên. Đối với các gói dữ liệu tiếp theo, giá trị nhãn thời gian tăng lên theo một đường thẳng tuỳ theo khoảng thời gian kể từ gói cuối cùng. Điều đó không đảm bảo truyền số liệu đến được đúng hạn, cung cấp các dữ liệu nhận được cho ứng dụng thu nhận với các đặc tính trễ giữa các gói khi rời khỏi trạm phát. Sự tồn tại của các bộ đệm với các thông tin mốc thời gian do RTP cung cấp sẽ cho phép người nhận tăng thêm trễ tới các gói, do đó làm quá trình truyền dữ liệu liên tục hơn.

2. Đánh số tuần tự (Sequence numbering). Hầu hết mạng gói cho phép các gói thuộc cùng 1 luồng dữ liệu đi qua mạng theo các đường dẫn khác nhau, mở rộng lượng truyền tải qua nhiều kết nối

khác nhau. Vì các kết nối này có thể có những đặc tính trễ khác nhau, rất có thể các gói sẽ đến người nhận theo một trật tự khác hẳn so với trật tự khi được truyền đi. Chuỗi RTP quy định một số ngẫu nhiên cho gói đầu tiên của một luồng và do đó gia tăng số lượng của mỗi một gói tiếp theo được gửi đi. Điều này cho phép người nhận quy định các gói trong 1 chuỗi đã định và phát hiện một hoặc một số lượng các gói thất thoát. Trong khi RTP không đòi hỏi khả năng tái truyền dẫn các gói bị mất, các ứng dụng nhận được thông báo về vị trí gián đoạn trong luồng audio.

3. Giám sát phân phối. RTP có thể cung cấp 1 phản hồi tới người gửi thông qua giao thức RTCP. RTCP xác định các gói tin về người gửi SR và người nhận RR bao gồm các thông tin chẳng hạn như jitter đến, số lượng gói bị mất, tổng số gói và octets đã được truyền đi và các số liệu khác. RTP cũng bỏ ngỏ khả năng các thông tin mào đầu RTP có thể bị ảnh hưởng của các nhà cung cấp dịch vụ mạng.

4. Xác định tải. Điều quan trọng không chỉ là dữ liệu audio được truyền qua RTP chính xác về đặc tính thời gian mà còn được giải mã tuỳ theo các quy định tương ứng. Do đó RTP xác định các thông tin nó truyền dẫn bằng việc quy định tải xác định cho từng gói. Các loại tải (đặc biệt là các luật mã hóa có thể được sử dụng cho audio và video) được định nghĩa trong RFC 1890.

Phương pháp tiếp cận của ITU đó là kết hợp RTP với các lớp ứng dụng mà nó hỗ trợ và không đứng đơn lẻ như một lớp giao thức riêng rẽ. Do đó RTP trở thành cơ sở giao thức để hỗ trở bất kể chức năng ứng dụng nào.

Định dạng tiêu đề gói RTP gồm 12 byte được mô tả trong hình 3.7 dưới đây.

V[2]	P	X	CC	M	PT	Số tuần tự
Nhãn thời gian						
Nhận dạng nguồn đồng bộ (SSRC)						
Nhận dạng nguồn (CSRC)						
Thoại						

Hình 3.7: Tiêu đề gói RTP

Các trường của tiêu đề:

- Trường *Version (V, 2 bit) (phiên bản)*: Cho biết phiên bản của RTP. Hiện tại sử dụng phiên bản 2 (theo RFC 1889).

- Trường *Padding (P, 1 bit) (đệm)*: Nếu bit này được thiết lập, gói tin chứa phần thông tin chèn thêm. Byte cuối cùng của phần chèn thêm cho biết số lượng byte được chèn thêm vào gói tin.

- Trường *Extension (X, 1 bit) (mở rộng):* bit mở rộng cho biết tiêu đề RTP có mở rộng thêm hay không. Tiêu đề RTP có chiều dài 4 byte bao gồm kiểu mở rộng (2 byte) và chiều dài mở rộng (2 byte). Tiêu đề mở rộng RTP phải được chèn sau trường trường có hiệu lực cuối cùng trong tiêu đề RTP chuẩn, đó là trường SSRC hoặc CSRC.

- Trường *CSRC Count (CC, 4 bit):* Số lượng các thành phần nhận dạng nguồn CSRC nằm trong phần mào đầu gói tin, số lượng này lớn hơn 1 nếu các gói tin RTP đến từ nhiều nguồn.

- Trường *Marker (M, 1 bit) (đánh dấu)*: dành cho lớp ứng dụng, chẳng hạn như khi đóng gói gói tin thoại sử dụng triệt khoảng lặng, bit này được dùng để đánh dấu gói tin bắt đầu có tín hiệu thoại tức là bắt đầu của một "talkspurt" (sau khoảng lặng). Đầu thu có thể sử dụng thông tin này để quyết định khi nào tạo nhiễu nền.

- Trường *Payload Type (PT, 7 bit) (kiểu tải tin):* Xác định loại tải trong gói RTP, dạng dữ liệu cũng như phương thức mã hóa/nén được sử dụng.

- Trường *Sequence Number (16 bit) (số tuần tự):* Giá trị này được tăng dần theo gói tin gửi đi. Nó được bên nhận dùng để sắp xếp các gói tin theo đúng thứ tự và phát hiện ra gói tin bị mất.

- Trường *Timestamp (32 bit) (nhãn thời gian):* Bên gửi thiết lập nhãn thời gian khi byte đầu tiên của gói tin được lấy mẫu và giá trị của nó được tăng dần. Nó được sử dụng để đồng bộ và tính toán jitter. Một gói chứa 20 ms tín hiệu thoại và được lấy mẫu ở tần số 8 kHz sẽ có nhãn thời gian tăng theo $8000 \times 0{,}02 = 160$.

- Trường *Synchronization Source Identifier (SSRC) (32 bit):* Được chọn ngẫu nhiên, được sử dụng để phân biệt giữa các nguồn phát trong cùng một phiên RTP. Nó xác định dữ liệu đến từ đâu (nếu chỉ có một nguồn phát) hoặc nó được ghép lại ở đâu (mixer) nếu có nhiều nguồn phát. Trường hợp có hai nguồn phát trong cùng phiên RTP có cùng SSRC, một cơ chế đơn giản trong RTP được sử dụng để giải quyết xung đột.

- Trường *Contributing Source Identifiers (CSRC) (32 bit):* 0 đến 15 giá trị CSRC, mỗi giá trị có kích thước 32 bít. Cho biết gói tin được tạo thành từ các nguồn thông tin nào.

Tiêu đề RTP/UDP/IP là 12, 8 và 20 bytes, tổng cộng tiêu đề của gói là 40 byte tức là gấp đôi phần tải tin khi sử dụng G.729 với hai mẫu thoại trong một gói (mỗi mẫu tương ứng với 10 ms nên một gói gồm hai mẫu là 20 ms). Tiêu đề RTP/UDP/IP có thể được nén sử dụng thuật toán nén nhãn tiêu đề RTP (cRTP). Cơ chế nén cRTP giảm tiêu đề RTP/UDP/IP từ 40 byte xuống còn 2 byte và xuống 4 byte khi UDP có sử dụng trường *checksum*.

Mặc dù một vài trường trong tiêu đề thay đổi giữa các gói RTP/UDP/IP tiếp theo. Bằng cách lưu giữ cả tiêu đề chưa nén và các giá trị của sai phân bậc nhất trong trạng thái phiên. cRTP cần thiết chỉ gửi sai phân bậc hai. Tiêu đề RTP đã nén được gửi vào vị trí của tiêu đề gói chưa nén khoảng 98% thời gian. Tiêu đề không nén phải được gửi theo chu kỳ để đảm bảo cả hai kết cuối có trạng thái đúng.

Các cơ chế trong RTP ở trên cung cấp các dịch vụ cần thiết để truyền tín hiệu thoại và video. Tuy nhiên, các kiểu mã hóa (codec) riêng biệt cũng yêu cầu thêm một số thông tin cần thiết để truyền đi. Vì mục đích này, RTP chấp nhận một số dạng tải tin *(payload)* khác nhau cho mỗi kiểu mã hóa riêng biệt. Các dạng tải tin này mô tả cú pháp và ý nghĩa của RTP tải tin và được biểu thị trong trường chỉ thị loại tải tin RTP. Trường 7 bit này được mô tả kiểu mã hóa (codec) thông qua việc gắn tên. Có 2 loại tải tin: loại tải tin tĩnh và tải tin động. Loại tải tin tĩnh được ấn định trước trong tài liệu RTP *profile*. Ví dụ, tải tin loại 0 dành cho mã hóa thoại luật μ.

3.4.2. Giao thức điều khiển RTP (RTCP)

Giao thức điều khiển RTP (mô tả trong RFC 1889) là RTCP. Các gói RTCP được sử dụng để cung cấp các thống kê về

chất lượng của phiên, thông tin người sử dụng và đồng bộ thời gian. RTCP đo chất lượng mạng bởi tính toán thống kê như trễ gói, mất gói và biến động trễ ở bộ thu. Sử dụng các thống kê này, kết cuối có thể thích nghi với các điều kiện biến động của mạng. Vì các gói RTCP góp phần tăng thêm nghẽn mạng, băng thông yêu cầu bởi RTCP là 5% tổng số băng thông phân bổ cho phiên. Hơn nữa, khoảng thời gian trung bình giữa các gói RTCP được đặt tối thiểu là 5 giây.

RTCP (*Real Time Control Protocol*) là giao thức điều khiển đồng hành cho RTP. Bên gửi tín hiệu (nguồn) và bên nhận gửi các gói RTCP theo chu kỳ tới cùng một nhóm đa hướng (nhưng các cổng khác nhau) giống như việc phân phối các gói RTP. Mỗi gói RTCP chứa một số phần tử, thường là báo cáo gửi *(SR- Sender Report)* hay báo cáo nhận *(RR- Receiver Report)* và tiếp theo là mô tả nguồn *(SDES - Source Descriptions)*. Mỗi phần tử thực hiện chức năng khác nhau.

Sender Report (SR) được các người sử dụng gửi đi (những người sử dụng này là các nguồn thông tin đa phương tiện được gửi đi trên RTP). Các phần tử thông tin này mô tả lượng thông tin đã gửi đi trước đó, cũng như là dữ liệu liên quan tới RTP nhãn thời gian và thời gian thực tế để đồng bộ giữa các luồng thông tin khác nhau.

Receiver Report (RR) được các bên tham gia phiên RTP (đang nhận luồng thông tin đa phương tiện) gửi đi. Mỗi một phần tử thông tin này chứa một khối cho một nguồn RTP trong nhóm. Mỗi khối mô tả tỷ lệ mất gói tích luỹ tại một thời điểm nào đó và sự trôi trượt từ nguồn đó. Thông tin trong khối cũng

chỉ ra nhãn thời gian cuối cùng và trễ khi nhận được báo cáo gửi, cho phép các bên gửi "nguồn" ước định được khoảng cách của chúng tới bên nhận.

Các gói *Source DEScriptor* (SDES) được sử dụng để điều khiển phiên. Chúng có chứa tên tuân thủ theo chuẩn gọi là CNAME (*Canonical Name*), dữ liệu nhận dạng duy nhất toàn cầu tương tự như địa chỉ email. CNAME được sử dụng để giải quyết tranh chấp về giá trị SSRC và kết hợp các luồng thông tin đa phương tiện khác nhau được gửi đi từ cùng một người sử dụng. Các gói SDES cũng dùng để nhận dạng thành viên tham gia thông qua tên, email và số điện thoại. Đây là một dạng đơn giản của việc điều khiển phiên. Các ứng dụng máy khách có thể hiển thị tên và địa chỉ email tại giao diện người sử dụng. Điều này cho phép các bên tham gia một phiên hiểu biết về các bên khác trong cùng một phiên. Nó cũng cho phép các bên có được thông tin liên lạc với nhau (như email hay số điện thoại).

Nếu một người sử dụng rời bỏ cuộc hội nghị, nó gửi bản tin *BYE*. Cuối cùng, phần tử *Application (APP)* có thể được sử dụng để bổ sung cho các gói RTCP thông tin đặc trưng cho ứng dụng.

Do các gói *báo cáo gửi*, *báo cáo nhận* và *SDES* chứa các thông tin thay đổi thường xuyên cho nên các gói này cần phải được gửi theo chu kỳ. Nếu các bên tham gia vào phiên RTP đơn giản chỉ cần gửi các gói RTCP theo chu kỳ cố định, băng thông sử dụng cho một nhóm truyền đa hướng sẽ tăng một cách tuyến tính với số thành viên trong nhóm. Điều này rõ ràng là không tốt. Thay vào đó, mỗi thành viên của phiên đếm số các thành viên khác trong phiên (thông qua các gói RTCP). Chu kỳ giữa các gói

RTCP từ mỗi một người sử dụng sau đó được thiết lập để phù hợp với số thành viên của nhóm. Điều này đảm bảo rằng băng thông sử dụng cho các gói RTCP vẫn không đổi, độc lập với số thành viên của nhóm. Do số thành viên trong nhóm được ước định nhờ việc đếm số các thành viên khác, mỗi thành viên mới phải mất ít nhiều thời gian để biết được số thành viên của nhóm.

3.5. ĐÁNH SỐ VÀ CHUYỂN ĐỔI ĐỊA CHỈ

3.5.1. Yêu cầu chung

ETSI đã đưa ra hai khuyến nghị TS 101 324 Ver.2.1.1 và TR 101 327 Ver.1.1.1 về yêu cầu đánh số đối với thuê bao VoIP để đảm bảo việc phối hợp hoạt động giữa hai mạng IP và mạng PSTN. Sau đây là một số yêu cầu chung:

- Mạng VoIP phải nhận dạng được số bị gọi theo chuẩn E.164 được sử dụng trong mạng quốc tế;

- Mạng VoIP có thể nhận dạng được mọi số bị gọi theo chuẩn E.164 được sử dụng trong mạng quốc gia;

- Mạng VoIP có thể nhận dạng được số bị gọi trong các mạng nội bộ;

- Mạng VoIP phải truyền được đầy đủ mọi tên khách hàng yêu cầu hạn chế nhận dạng thuê bao chủ gọi;

- Mạng VoIP phải hỗ trợ cơ chế lựa chọn nhà cung cấp mạng theo quy định của từng quốc gia trong trường hợp có nhiều nhà cung cấp mạng;

Chú ý: Việc lựa chọn nhà cung cấp mạng có thể được thực hiện bằng nhiều cách: đặt ngầm định, lựa chọn trước bởi người sử

dụng, quay số mã truy nhập và mã nhận dạng của nhà cung cấp mạng, hoặc bằng một cách thức khác do quốc gia đó quy định.

- Mạng VoIP phải hỗ trợ chức năng phân tích số để lựa chọn nhà cung cấp dịch vụ nếu cần thiết;

Chú ý: Các quốc gia khác nhau có yêu cầu khác nhau về số để lựa chọn nhà cung cấp dịch vụ. Một vài quốc gia yêu cầu chức năng bổ sung trong mạng khởi phát cuộc gọi và mạng chuyển tiếp.

3.5.1.1 Yêu cầu đối với cuộc gọi từ mạng IP đến PSTN

Trong trường hợp này (trường hợp 1), mạng VoIP phải cung cấp các chức năng sau:

- Thuê bao chủ gọi trong mạng VoIP phải nhận dạng được số thuê bao bị gọi kiểu E.164 trong mạng PSTN, hoặc số sử dụng trong mạng chuyển mạch kênh nội bộ trong trường hợp mạng IP nội bộ kết nối với mạng chuyển mạch kênh nội bộ;

- Mạng VoIP phải có khả năng cung cấp nhận dạng thuê bao chủ gọi (CLI) cho mạng PSTN nếu có yêu cầu ngay cả khi thuê bao chủ gọi kích hoạt dịch vụ hạn chế CLI. Các thông tin CLI phải nhận dạng được bởi hoặc đầu cuối hoặc người sử dụng hoặc cả hai;

Chú ý: Các thông tin CLI được mạng yêu cầu để phục vụ mục đích quản lý thậm chí kể cả khi nó không được chuyển đến thuê bao bị gọi.

- Trong trường hợp mạng có sử dụng các mã tiền tố (ví dụ như mã vùng), thì mạng VoIP cũng phải hỗ trợ mã tiền tố;

- Trong trường hợp mạng có sử dụng các số dịch vụ đặc biệt (ví dụ như 116,113, 115...) thì mạng VoIP cũng phải hỗ trợ sử dụng các số này.

3.5.1.2 Yêu cầu đối với cuộc gọi từ mạng PSTN đến IP

Trong trường hợp này (trường hợp 2), mạng VoIP phải cung cấp các chức năng sau:

- Mạng VoIP phải xác nhận đối với mọi yêu cầu hạn chế nhận dạng thuê bao chủ gọi nhận được từ mạng PSTN;

- Trong trường hợp có nhiều nhà cung cấp dịch vụ thì mạng VoIP phải hỗ trợ việc lựa chọn này.

3.5.1.3 Yêu cầu đối với cuộc gọi từ PSTN→ IP→ PSTN

Trong trường hợp này (trường hợp 3), mạng VoIP phải cung cấp các chức năng sau:

- Mạng VoIP phải truyền được các thông tin CLI, các thông tin báo hiệu tương ứng được cung cấp bởi mạng PSTN khởi phát và các số thuê bao bị gọi để có thể thiết lập thành công cuộc gọi.

Chú ý: Trong trường hợp này mạng IP chỉ cung cấp kết nối giữa các mạng PSTN, vì vậy số thuê bao chủ gọi và bị gọi chính là số của chúng trong mạng PSTN và chúng có thể không có địa chỉ IP tương ứng. Vì vậy việc định tuyến cuộc gọi trong mạng IP sẽ dựa trên cuộc gọi đó xuất phát từ đâu trong mạng PSTN và các phần của số thuê bao bị gọi theo định dạng E.164 ví dụ như mã quốc gia, mã vùng hoặc mã nút của mạng nội bộ để định tuyến cuộc gọi đến đích trong mạng PSTN.

3.5.1.4 Yêu cầu đối với cuộc gọi từ IP → PSTN → IP

Trong trường hợp này (trường hợp 4), mạng VoIP phải cung cấp các chức năng như trong các trường hợp 1 và 2.

3.5.1.5 Các phương thức quay số

Phương thức thuê bao chủ gọi quay số để thiết lập cuộc gọi với thuê bao bị gọi được chia làm hai loại: phương thức thiết lập một giai đoạn và phương thức thiết lập hai giai đoạn.

- Phương thức thiết lập một giai đoạn: thuê bao chủ gọi có thể kết nối trực tiếp với thuê bao bị gọi bằng cách bấm số với chỉ một chuỗi số liên tiếp là số điện thoại của thuê bao bị gọi. Chuỗi số được bấm ở loại này có thể bao gồm một tiền tố và mã của nhà cung cấp mạng hoặc nhà cung cấp dịch vụ.

- Phương thức thiết lập hai giai đoạn: thuê bao chủ gọi trước tiên phải kết nối với một trạm trung gian ví dụ như điểm truy nhập của nhà cung cấp dịch vụ hoặc PBX đặt tại nhà nơi thuê bao bị gọi bằng cách bấm số điện thoại của trạm đó. Sau khi đã kết nối được với trạm trung gian, thuê bao chủ gọi phải bấm tiếp số điện thoại của thuê bao bị gọi để thiết lập cuộc gọi. Thông thường lần bấm số thứ hai này xảy ra sau khi thuê bao chủ gọi nhận được tín hiệu phúc đáp từ trạm trung gian.

Đối với người sử dụng, phương thức thiết lập một giai đoạn thông dụng và thuận tiện hơn bởi vì số lượng chữ số người sử dụng phải bấm ít hơn so với phương thức hai giai đoạn. Bên cạnh đó, nhà cung cấp mạng và nhà cung cấp dịch vụ cũng thích phương thức thiết lập một giai đoạn hơn do thói quen của người sử dụng.

3.5.1.6 Các số lựa chọn

Năm 1999, ITU-T đã nghiên cứu về phương thức và cách đánh các số lựa chọn trong mạng chuyển mạch kênh và đưa ra tiêu chuẩn chuẩn quốc tế cho các số này. Theo ITU-T, các số lựa chọn bao gồm 3 loại sau:

- Lựa chọn vùng: cho phép thuê bao của mạng chuyển mạch kênh giữ nguyên số theo khuyến nghị E.164 khi họ di chuyển từ vùng này sang vùng khác.

- Lựa chọn nhà cung cấp dịch vụ: cho phép thuê bao của mạng chuyển mạch kênh giữ nguyên số theo khuyến nghị E.164 khi người đó chuyển từ sử dụng dịch vụ của nhà cung cấp dịch vụ này sang nhà cung cấp dịch vụ khác.

- Số lựa chọn dịch vụ: cho phép thuê bao của mạng chuyển mạch kênh giữ nguyên số theo khuyến nghị E.164 khi người đó chuyển từ sử dụng dịch vụ này sang dịch vụ khác.

Trong các khuyến nghị của ITU-T, các số lựa chọn được giả sử là có thể thực hiện được giữa các số theo khuyến nghị E.164 trong cùng một quốc gia.

Mặc dù việc nghiên cứu về các số lựa chọn chủ yếu được ứng dụng trong mạng chuyển mạch kênh, tuy nhiên khái niệm về số lựa chọn còn có thể được áp dụng trong các dịch vụ viễn thông khác bao gồm cả dịch vụ điện thoại IP. IETE và ETSI đã đưa ra các khuyến nghị riêng biệt về khái niệm đánh địa chỉ linh hoạt trong dịch vụ điện thoại IP.

3.5.2. Phương pháp đánh số thuê bao

3.5.2.1 Yêu cầu đối với quy tắc đánh số

Quy tắc đánh số trong mạng VoIP được trình bày trong khuyến nghị TS 101 324 Ver.2.1.1 và TR 101 327 Ver.1.1.1 của ETSI phải đáp ứng các yêu cầu tối thiểu sau đây:

- Các số có thể chỉ bao gồm các số thập phân;

- Độ dài của số thuê bao có thể được sử dụng trong mạng toàn cầu hoặc trong việc phối hợp giữa các mạng nội bộ;

- Số thuê bao phải là duy nhất trong phạm vi toàn cầu đối với mạng công cộng;

- Các số thuê bao phải cho phép người sử dụng quay số một cấp;

- Các số phải hỗ trợ việc di động trong một vùng và liên vùng theo khuyến nghị TR 101 338;

- Các số phải hỗ trợ lựa chọn nhà cung cấp dịch vụ trong khi vẫn giữ nguyên số.

3.5.2.2 Quy tắc đánh số để hỗ trợ tại giao diện đối với mạng PSTN

Quy tắc sau được áp dụng đối với mọi trường hợp cuộc gọi.

a) Đối với mạng VoIP công cộng

Trong mạng VoIP công cộng, các số địa chỉ trao đổi với mạng PSTN phải tuân theo khuyến nghị E.164 của ITU-T.

Chú ý: khuyến nghị E.164 của ITU-T đưa ra nhiều cách lựa chọn đánh số khác nhau. Có thể cùng một lúc lựa chọn một hay nhiều khuôn dạng đánh số này. Đối với mạng VoIP không cần thiết phải đánh số tương ứng một-một giữa một số E.164 đối với một đầu cuối trong mạng VoIP vì có thể có nhiều người sử dụng cùng sử dụng một đầu cuối VoIP.

b) Đối với mạng VoIP nội bộ

Trong mạng VoIP nội bộ, các địa chỉ trao đổi với mạch chuyển mạch kênh nội bộ phải tuân theo khuyến nghị ETS 300 189 hoặc ISO/IEC 11571.

Trong mạng VoIP nội bộ, các địa chỉ trao đổi với mạng PSTN phải tuân theo khuyến nghị E.164 của ITU-T.

3.5.2.3 Phương pháp đánh số thuê bao

a) Quy tắc của IETF

Theo IETF, hệ thống đánh số cho điện thoại IP dựa trên nguyên tắc mỗi số E.164 được đăng kí cho mỗi người sử dụng để có thể truy nhập Internet thông qua bất cứ một đầu cuối nào của mạng. Năm 1998 Lee và Orsis đã đưa ra hai ưu điểm của việc ấn định số E.164 đối với người sử dụng máy tính truy cập Internet. Việc ấn định số E.164 cho phép người sử dụng dễ dàng hơn trong việc sử dụng cổng điện thoại để gọi cho người khác sử dụng máy tính. Bên cạnh đó, những chức năng sẵn có trên mạng Internet có thể được sử dụng để cung cấp cho dịch vụ điện thoại IP.

Khi việc định tuyến trên Internet dựa trên số E.164 đã đăng ký ứng với thuê bao bị gọi, thì số E.164 cần phải chuyển đổi thành địa chỉ IP tương ứng. Việc chuyển đổi này được thực hiện thông qua cơ sở dữ liệu của mạng được đặt phân tán tại mỗi cổng kết nối của nhà cung cấp dịch vụ. Có 2 phương án đối với cấu trúc hình cây có ưu tiên của cơ sở dữ liệu để chuyển đổi giữa số E.164 và địa chỉ IP.

** Sử dụng hệ thống tên theo vùng cho điện thoại IP*

Năm 1998 Faltstrom đã đưa ra nguyên tắc sử dụng hệ thống tên vùng (DNS) cho dịch vụ điện thoại IP. DNS cung cấp chuyển đổi giữa tên vùng và địa chỉ IP. Khi số E.164 được ghi theo dạng tên vùng thì DNS có thể thực hiện chuyển đổi từ số E.164 và địa chỉ IP cho dịch vụ điện thoại IP.

Nguyên tắc: tạo một miền phụ e164.int và ghi lại số E.164 dưới dạng tên vùng. Trong cấu trúc số E.164, đầu tiên là mã

quốc gia (CC), sau đó là mã nước của thuê bao bị gọi (NDC) rồi đến số của thuê bao bị gọi (SN). Mặt khác, cấu trúc của tên vùng là: đầu tiên là vùng mức thấp, sau đó là vùng mức cao. Chính vì vậy tên vùng của số E.164 có cấu trúc đối lập với số E.164 ban đầu và viết theo thứ tự ngược lại.

Ví dụ: Số E.164 là: "+ 1 555 123 4567"

Số tên vùng của nó là : "7.6.5.4.3.2.1.5.5.5.1.e164.int".

Ứng dụng nguyên tắc trên bằng cách sử dụng phương thức vùng phụ tcp.int cho dịch vụ điện thoại IP. Năm 1998, Brown đã đề xuất cấu trúc cơ sở dữ liệu của số E.164 cho phép ánh xạ trong cấu trúc hình cây. Trong cấu trúc hình cây của một cơ sở dữ liệu, mỗi chữ số của số E.164 đóng vai trò như một điểm nút. Cơ sở dữ liệu này đem lại nhiều lợi thế trong việc triển khai các dịch vụ trong tương lai.

b) Khuyến nghị của ETSI

ETSI đã nghiên cứu về ứng dụng của điện thoại IP và tên dự án của nó là TIPHON. Những yêu cầu đối với loại hình kết nối PC - điện thoại:

- Thuê bao bị gọi sử dụng điện thoại đã phải đăng kí số E.164 trong mạng chuyển mạch kênh.

- Thuê bao chủ gọi sử dụng máy tính thông qua Internet có thể quay số E.164 để xác định thuê bao bị gọi. Số E.164 sẽ được làm rõ phần nào là địa chỉ cổng, phần nào là địa chỉ của đầu cuối hoặc người sử dụng.

- Việc chuyển đổi từ số E.164 thành dạng địa chỉ IP cần phải được sẵn sàng, số E.164 được sử dụng để nhận ra được thuê bao bị gọi và thiết lập đường nối từ PC đến điện thoại.

Vì thế số E.164 được quy định là thông tin địa chỉ được chuyển thông qua mạng chuyển mạch kênh và Internet.

Bên cạnh đó trong dịch vụ điện thoại IP cũng yêu cầu nhận dạng được thuê bao chủ gọi, nhận dạng được loại cuộc gọi như những cuộc gọi cấp cứu khẩn cấp v.v... Vì vậy, mạng chuyển mạch kênh cũng yêu cầu thuê bao chủ gọi phải đăng kí số E.164 vì chỉ có số E.164 mới có thể được sử dụng để nhận dạng thuê bao chủ gọi trong mạng.

ETSI đã đưa ra 3 kiểu cấu trúc của số E.164 nên sử dụng trong dịch vụ điện thoại IP. Đó là: số định vị theo vùng địa lý quốc gia, số cho các quốc gia không chia vùng địa lý và số định vị toàn cầu.

3.5.3. Phương pháp chuyển đổi số E.164 và địa chỉ IP

3.5.3.1 Khuyến nghị của IETF

Năm 1998, Rosenberg và Schulzrinne đã đưa ra cơ chế chuyển đổi giữa số E.164 và địa chỉ IP trong trường hợp kết nối điện thoại - điện thoại và PC - điện thoại. Họ đã đưa ra các đặc tính cơ bản của cổng kết nối cho dịch vụ điện thoại IP và đề xuất các máy chủ định vị. Mỗi cổng kết nối được xác định bởi 3 thông số:

- Vùng số E.164 mà nó có thể cung cấp dịch vụ.
- Số lượng dịch vụ mà nó có thể cung cấp được.
- Kiểu dịch vụ mà nó có thể cung cấp.

Những đặc điểm này được sử dụng để máy chủ lựa chọn cổng kết nối và thiết lập đường truyền cho cuộc gọi.

Năm 1998, Agapi và một số người khác đã đưa ra cơ cấu chuyển đổi địa chỉ với loại hình kết nối máy tính - điện thoại và

máy tính - máy tính và đưa ra 3 mô hình kinh doanh khi sử dụng phối hợp hoạt động với máy chủ định vị như sau:

- ITSP (nhà cung cấp dịch vụ thoại Internet) triển khai và duy trì hoạt động của tất cả các cổng kết nối và máy chủ định vị. Máy chủ định vị của ITSP chịu trách nhiệm chuyển đổi số E.164 thành địa chỉ IP của cổng kết nối thích hợp để thiết lập cuộc thoại.

- Các ITSP phối hợp với nhau chia sẻ các cổng kết nối và thông tin địa chỉ của các cổng kết nối đó với nhau. Máy chủ định vị của các ITSP trao đổi thông tin kết nối với nhau. Các máy chủ định vị thực hiện chức năng chuyển đổi địa chỉ E.164 của thuê bao bị gọi thành địa chỉ IP của cổng kết nối dùng chung trong mạng, xác định cổng kết nối thích hợp nhất để thiết lập cuộc thoại.

- Tất cả các ITSP có thể chia sẻ cổng kết nối và thông tin địa chỉ cổng kết nối trong mạng công cộng. Tất cả các máy chủ định vị trao đổi thông tin của chúng với nhau và chia sẻ các thông tin chuyển đổi giữa số E.164 của thuê bao bị gọi và địa chỉ IP của cổng kết nối.

- Năm 1998 Lee và Orsic đã đưa ra phương thức chuyển đổi địa chỉ trong loại hình kết nối điện thoại - máy tính, máy tính - máy tính. Ở hai trường hợp này, thuê bao bị gọi là máy tính trên Internet. Giả sử rằng mỗi khách hàng sử dụng PC kết nối Internet có một số E.164 và mỗi kết nối Internet này có một địa chỉ IP. Mô hình chuyển đổi này gần giống với mô hình do Rosenberg và Schulzrinne đưa ra.

3.5.3.2 Mô hình phối hợp hoạt động trong chuyển đổi địa chỉ

Các khuyến nghị của Agapi và các nhà nghiên cứu khác trong việc chuyển đổi địa chỉ có thể được mở rộng và áp dụng

cho tất cả các loại kết nối điện thoại IP. Sau đây là 3 mô hình tổng quát: mạng riêng, mạng quan hệ mật thiết và mạng liên kết mở. Sau đó phần này sẽ đưa ra phương thức định tuyến trên ba mô hình này.

** Mô hình mạng riêng*

Mỗi ITSP tự triển khai và duy trì hoạt động của tất cả các cổng kết nối cần thiết và các máy chủ để cung cấp dịch vụ điện thoại IP. ITSP lưu trữ các thông tin địa chỉ cổng kết nối của họ trong các máy chủ của họ. Nếu các ITSP khác thiết lập cuộc thoại IP thì nó cũng lưu giữ các thông tin địa chỉ vào ô nhớ trong máy chủ ứng với thuê bao của họ tham gia cuộc thoại để phục vụ cho các lần kết nối sau.

** Mô hình nhóm quan hệ mật thiết*

Một vài ITSP chia sẻ cổng kết nối khi liên kết. Mỗi ITSP trong liên kết lưu trữ thông tin địa chỉ cổng kết nối trong máy chủ của nó. Ngoài ta, nó còn có thể cung cấp thông tin địa chỉ của thuê bao khi thiết lập dịch vụ. Chính vì thế các ITSP có thể chia sẻ thông tin địa chỉ và do đó cần phải có một chuẩn chung cho dữ liệu địa chỉ trong các liên kết.

** Mô hình liên kết mở*

Tất cả các ITSP chia sẻ cổng kết nối khi liên kết. Mỗi ITSP trong liên kết lưu trữ thông tin địa chỉ cổng trong máy chủ của nó. Ngoài ta, nó còn có thể cung cấp thông tin địa chỉ của thuê bao khi thiết lập dịch vụ. Cấu trúc hệ thống dữ liệu mở được yêu cầu để chia sẻ các thông tin địa chỉ.

ITSP cung cấp dịch vụ chuyển tiếp theo các thông tin địa chỉ chia sẻ. Khi một thuê bao của mạng chuyển mạch kênh bị gọi thì máy chủ của ITSP sẽ chuyển đổi số E.164 thành địa chỉ IP của cổng kết nối phù hợp nhất. Khi một thuê bao máy tính kết nối Internet bị gọi thì máy chủ sẽ chuyển đổi số E.164 thành địa IP của thuê bao đó.

Mỗi ITSP có thể lựa chọn mô hình thích hợp cho mục đích kinh doanh của mình. Tuy nhiên mỗi ITSP chỉ có thể chọn một trong ba mô hình, không thể chọn 2 mô hình hoặc nhiều hơn tại cùng một thời điểm. Việc chuyển đổi thông tin địa chỉ giữa các ITSP cần phải hỗ trợ lẫn nhau.

Ở mô hình liên kết mạng riêng, do tất cả các hoạt động đòi hỏi chuyển đổi địa chỉ đều do ITSP đó tự thực hiện nên ITSP có thể sử dụng những giao thức riêng cho việc chuyển đổi. Tuy nhiên trong mô hình liên kết mạng riêng có những hạn chế về cung cấp dịch vụ do sự giới hạn của các cổng kết nối. Bên cạnh đó, nó không cung cấp được các dịch vụ chuyển tiếp khi thuê bao bị gọi là máy tính sử dụng Internet đăng ký dịch vụ với một ITSP khác.

Ở mô hình nhóm quan hệ mật thiết, ITSP cần sử dụng giao thức thông thường cho việc chuyển đổi. Giao thức này có thể độc lập với các liên kết. Do các cổng kết nối được chia sẻ với các ITSP khác trong liên kết nên mô hình nhóm quan hệ mật thiết sẽ có ít hạn chế hơn trong việc cung cấp dịch vụ so với mô hình liên kết mạng riêng. Tuy nhiên giống mô hình liên kết mạng riêng, nó không cung cấp được các dịch vụ chuyển tiếp khi thuê bao bị gọi là máy tính sử dụng Internet đăng ký dịch vụ với các ITSP không nằm trong liên kết.

Bên cạnh đó, đối với hệ thống liên kết mở, việc liên kết giữa các cổng kết nối có thể không giới hạn các dịch vụ mà ITSP cung cấp. Tuy nhiên, ITSP cần sử dụng giao thức chuẩn cho việc chuyển đổi.

* Định tuyến cho các loại hình dịch vụ

- Định tuyến trong kết nối điện thoại - điện thoại

Trong loại hình liên kết này, thuê bao chủ gọi lựa chọn ITSP có thể cung cấp dịch vụ chuyển giao cho họ. ITSP được chọn sẽ lựa chọn một cổng riêng để thiết lập đường dẫn tới thuê bao bị gọi.

Nếu ITSP được lựa chọn theo môt hình liên kết mạng riêng, nó cần lựa chọn cổng. Nếu số lượng cổng có thể chia sẻ càng nhiều thì khả năng lựa chọn càng cao. Khi đó quãng đường truy nhập vào mạng chuyển mạch kênh sẽ ngắn hơn hoặc ta sẽ có một vùng cung cấp dịch vụ lớn hơn. Với loại hình liên kết theo hệ thống mở có thể sẽ mang lại cho ITSP một khả năng lựa chọn lớn nhất cho cổng riêng biệt và một khoảng truy nhập ngắn nhất trong mạng chuyển mạch kênh.

- Định tuyến trong kết nối điện thoại - máy tính

Trong loại hình liên kết này, ITSP cung cấp việc thiết lập dịch vụ cho thuê bao bị gọi. Thuê bao chủ gọi sử dụng điện thoại mạng chuyển mạch kênh có thể không biết thuê bao bị gọi sử dụng máy tính trên Internet. Thuê bao chủ gọi không có ý định sử dụng dịch vụ chuyển giao trong loại hình dịch vụ điện thoại IP. Khi đó, thuê bao chủ gọi chỉ bấm số E.164 của thuê bao bị gọi. Do vậy, trong cấu trúc của số E.164 của thuê bao bị gọi cần

phải có một phần đặc biệt để bộ phận định tuyến trong mạng chuyển mạch kênh có thể nhận ra thuê bao bị gọi sử dụng máy tính trên Internet và truy cập vào cổng của ITSP. Mã nhận dạng (IC) hoặc một vài chữ số của mã nước bị gọi (NDC) có thể sử dụng như là một phần của số E.164 cho việc thông báo dịch vụ điện thoại IP. Điều này sẽ gây thêm sự phức tạp trong hệ thống định tuyến trong mạng chuyển mạch kênh.

- Định tuyến trong kết nối từ máy tính - điện thoại

Trong loại hình kết nối từ máy tính đến điện thoại, người gọi lựa chọn ITSP có khả năng cung cấp dịch vụ chuyển giao cho họ. ITSP cung cấp dịch vụ chuyển giao cho thuê bao chủ gọi cần lựa chọn cổng thích hợp để thiết lập đường dẫn đến thuê bao bị gọi thông qua nó. Các phương thức khác nhau đưa ra các đường dẫn khác nhau.

Nếu ITSP được lựa chọn theo môt hình liên kết mạng riêng, nó cần lựa chọn cổng. Nếu số lượng cổng có thể chia sẻ càng nhiều thì khả năng lựa chọn càng cao. Khi đó quãng đường truy nhập vào mạng chuyển mạch kênh sẽ ngắn hơn hoặc ta sẽ có một vùng cung cấp dịch vụ lớn hơn. Với loại hình liên kết theo hệ thống mở có thể sẽ mang lại cho ITSP một khả năng lựa chọn lớn nhất cho cổng riêng biệt và một khoảng truy nhập ngắn nhất trong mạng chuyển mạch kênh.

- Định tuyến trong kết nối máy tính - máy tính

Trong loại hình kết nối máy tính - máy tính, thuê bao chủ gọi lựa chọn ITSP có khả năng cung cấp dịch vụ chuyển giao cho họ. Bên cạnh đó, ITSP cung cấp dịch vụ đầu cuối cho thuê

bao bị gọi. ITSP được lựa chọn bởi thuê bao chủ gọi cần tìm ra địa chỉ IP của thuê bao bị gọi thông qua số E.164 được bấm bởi thuê bao chủ gọi. Cần phải lưu ý rằng ITSP được lựa chọn bởi thuê bao chủ gọi cần phải khác với ITSP cung cấp dịch vụ đầu cuối cho thuê bao bị gọi. Vì thế những gì xảy ra với việc định tuyến trong loại hình kết nối giữa máy tính với máy tính cũng giống như trong loại hình điện thoại với máy tính khi thuê bao chủ gọi có ý định sử dụng dịch vụ chuyển giao của dịch vụ điện thoại IP.

Nếu ITSP được lựa chọn bởi thuê bao chủ gọi cung cấp những dịch vụ đầu cuối cho thuê bao bị gọi, ITSP có thể khôi phục lại được địa chỉ IP của thuê bao bị gọi từ cơ sở dữ liệu của nó bất kể mô hình liên kết của ITSP. Vì thế ITSP có thể thiết lập đường dẫn từ thuê bao chủ gọi đến thuê bao bị gọi.

Mặt khác ITSP được lựa chọn từ thuê bao chủ gọi khác với ITSP cung cấp thiết dịch vụ đầu cuối cho thuê bao bị gọi.

Với ITSP hoạt động theo mô hình liên kết mạng riêng biệt, ITSP có thể cung cấp dịch vụ chuyển giao cho thuê bao chủ gọi lại không thể truy nhập cơ sở dữ liệu theo thông tin địa chỉ của thuê bao bị gọi đã được lưu trữ và do đó đường dẫn không được thiết lập.

Với ITSP hoạt động theo mô hình liên kết đóng gói và những ITSP khác không có cùng mối liên kết, ITSP không thể truy nhập cơ sở dữ liệu theo thông tin địa chỉ của thuê bao bị gọi đã được lưu trữ và do đó đường dẫn không được thiết lập.

Nếu cả hai ITSP hoạt động theo mô hình liên kết đóng gói trong cùng một mối liên kết hoặc nếu cả hai ITSP hoạt động theo

mô hình liên kết hệ thống mở, ITSP có thể khôi phục lại được địa chỉ IP của thuê bao bị gọi từ cơ sở dữ liệu đã chia sẻ.

Giống như biểu hiện của việc định tuyến trong kết nối điện thoại - máy tính, khi một thuê bao chủ gọi có ý định sử dụng dịch vụ chuyển giao của điện thoại IP, một phần đặc biệt trong số E.164 của thuê bao chủ gọi thông báo dịch vụ điện thoại IP không được sử dụng cho việc định tuyến trong trường này.

3.5.4. Phương pháp định tuyến giữa PSTN và IP

Khi thực hiện một cuộc gọi mà thuê bao chủ gọi và thuê bao bị gọi nằm riêng rẽ trong mạng chuyển mạch kênh và Internet, thuê bao chủ gọi phải đưa ra thông tin địa chỉ của thuê bao bị gọi để hệ thống định tuyến trong mạng chuyển mạch kênh và mạng Internet căn cứ vào đó định tuyến cuộc gọi. Có 3 cách khác nhau để định tuyến cuộc thoại IP giữa các thuê bao nằm riêng rẽ trong mạng chuyển mạch kênh và Internet.

3.5.4.1 Cách thứ nhất

Thêm các chức năng mới cho mạng chuyển mạch kênh để xử lý các thông tin định tuyến sử dụng trong mạng Internet. Các thông tin địa chỉ được sử dụng trên Internet có thể là địa chỉ IP, tên vùng, địa chỉ e-mail v.v... Những người sử dụng dịch vụ điện thoại IP cần được đăng kí địa chỉ IP. Thuê bao chủ gọi sẽ chỉ định thuê bao bị gọi thông qua các thông tin địa chỉ của thuê bao bị gọi. Cấu trúc của nó giống như cấu trúc tên vùng bằng việc sử dụng các chữ số từ 0 - 9 và các kí hiệu # và *. Do đó cần phải đưa ra các quy ước khi sử dụng những kí tự này. Tuy nhiên điều này có thể không khả thi do việc đưa địa chỉ vào sẽ rất phức tạp.

3.5.4.2 Cách thứ hai

Thêm các chức năng mới cho mạng chuyển mạch kênh để giao tiếp với số E.164 trên Internet. Mỗi thuê bao sử dụng dịch vụ điện thoại IP được đăng kí số E.164. Mỗi chức năng mới ra đời cần đưa ra phép chuyển đổi từ số E.164 sang địa chỉ IP. Khi đó người sử dụng chỉ đưa vào số E.164 và hệ thống coi đó như là thông tin địa chỉ của người sử dụng dịch vụ điện thoại IP. Tuy nhiên do dịch vụ điện thoại IP sử dụng chung số E.164 với các dịch vụ khác nên dung lượng người sử dụng dịch vụ điện thoại IP sẽ bị giới hạn.

3.5.4.3 Cách thứ ba

Thêm các cách thức đánh địa chỉ mới độc lập với hệ thống địa chỉ IP và số E.164. Một hệ thống đánh địa chỉ mới cần được sử dụng cho dịch vụ điện thoại IP. Hệ thống này cần phải cải tiến từ các hệ thống đánh địa chỉ cũ và áp dụng cho cả thuê bao chủ gọi và thuê bao bị gọi.

Tuy nhiên cũng giống như hệ thống đánh địa chỉ IP, điều này có thể không khả thi do việc đưa địa chỉ vào sẽ rất phức tạp. Vì vậy tác giả quyết định lựa chọn phương thức thứ hai là phương thức đăng kí địa chỉ cho các thuê bao sử dụng dịch vụ điện thoại IP.

3.5.5. Khuyến nghị

Giải pháp phù hợp nhất đối với hệ thống đánh số trong mạng điện thoại qua Internet bao gồm ba yếu tố sau đây:

- Đánh số E.164 toàn cầu cho người sử dụng theo máy tính đầu cuối trên Internet.

- Phương pháp hệ thống mở đối với cơ chế lập sơ đồ giữa số E.164 và các địa chỉ IP3.

- Yêu cầu đối với trợ giúp cho quay số một giai đoạn không qua mã truy nhập nhà cung cấp dịch vụ.

Sự kết hợp giữa các yếu tố này đem lại dung lượng đánh số lớn cho người sử dụng trên Internet, cho phép sắp xếp dễ dàng các ứng dụng di chuyển số và tránh trường hợp đường dẫn không thể thiết lập.

Cơ quan có thẩm quyền nên quyết định cách thức và mức độ mà điện thoại qua Internet được coi là dịch vụ điện thoại. Cơ quan có thẩm quyền cần phải xem xét các khía cạnh sau đây khi đưa ra chính sách:

- Hệ thống phân số E.164 cần phải thay đổi để đồng thời phân cho người sử dụng trên Internet.

- Cần có quy tắc phối hợp giữa nhà cung cấp dịch vụ VoIP (ITSP) và nhà cung cấp dịch vụ trên mạng chuyển mạch kênh cho điện thoại qua Internet.

- Ứng dụng di chuyển số cho điện thoại qua Internet có thể đòi hỏi việc củng cố các ITSP và nhà cung cấp dịch vụ trên mạng chuyển mạch kênh.

Cơ sở dữ liệu cho dịch vụ điện thoại qua Internet được chia thành cơ sở dữ liệu hành chính gồm thông tin phân địa chỉ và cơ sở dữ liệu sử dụng thời gian thực để định tuyến. Trong khi cơ sở dữ liệu hành chính cần phải được quản lý dưới quyền phân thông tin địa chỉ, việc quản lý cơ sở dữ liệu sử dụng thời gian thực có thể không cần phụ thuộc vào quyền này. Tuy nhiên, giao thức

chuẩn mở là cần thiết để trao đổi dữ liệu lập sơ đồ giữa các cơ sở dữ liệu sử dụng thời gian thực với nhau.

Việc phân số E.164 cho người sử dụng trên Internet được coi là đòi hỏi tối thiểu đối với mối quan hệ qua lại giữa mạng chuyển mạch kênh và Internet. Những hệ thống địa chỉ khác với hệ thống E.164 có thể được sử dụng trong kết nối máy tính đến máy tính. Hệ thống phân địa chỉ phù hợp khác cho kết nối máy tính đến máy tính cần phải được nghiên cứu thêm.

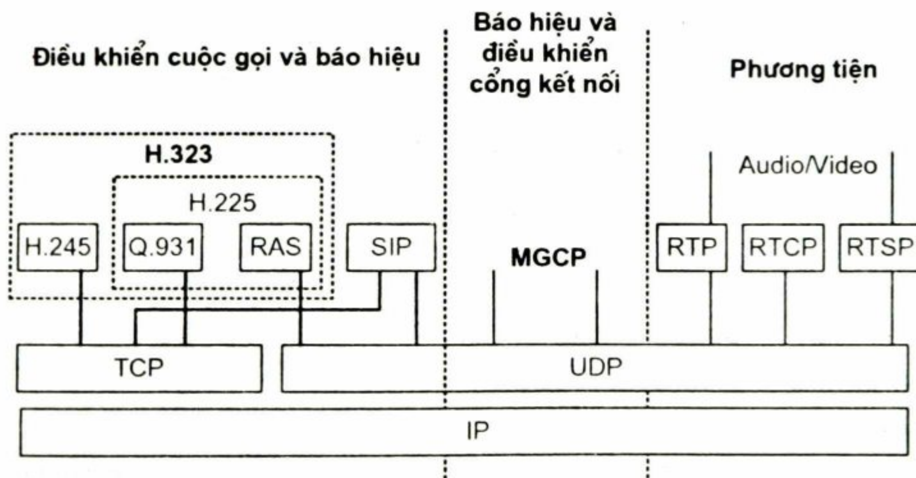
Chương IV
CÁC GIAO THỨC ĐIỀU KHIỂN

4.1. GIỚI THIỆU

So với thoại truyền thống, thoại IP hoạt động trong một môi trường mới, hiện đã có một số các giao thức lớp ứng dụng phát triển cho thoại Internet. Một điểm quan trọng để có một hạ tầng cơ sở gói dựa trên tiêu chuẩn là khả năng có các tiêu chuẩn mở cho các lớp điều khiển cuộc gọi. Những chuẩn mở này được cung cấp bởi các giao thức như H.323, MGCP, SIP và các giao diện mở được phát triển rộng khắp qua cơ sở hạ tầng đóng gói. Một trong các công việc của giao thức điều khiển cuộc gọi là chỉ cho các dòng RTP bắt đầu ở đâu và kết thúc ở đâu. Giao thức điều khiển cuộc gọi hoàn tất nhiệm vụ này bằng việc biên dịch giữa các địa chỉ IP và các số điện thoại.

Hiện nay chúng ta chưa thể thay thế hoàn toàn mạng điện thoại chuyển mạch công cộng (PSTN) bằng công nghệ VoIP bởi còn nhiều điều phức tạp bên trong các giao thức mới chạy trên nền tảng của IP. Tuy nhiên, gần đây các giao thức cho báo hiệu cuộc gọi và điều khiển thiết bị đang được chuẩn hóa, chúng ta đang gần đạt đến một môi trường có tính liên kết hoạt động cao. Giao thức điều khiển truyền thông (MGCP) hiện nay đã trở thành tiêu chuẩn chính thức, trong khi đó các cải tiến được thừa

nhận gần đây trong phiên bản 4 của H.323 đã tạo điều kiện thuận lợi khi kết hợp với các giao thức khác để tạo ra các giải pháp cho hệ thống truyền thoại hoàn chỉnh và đặc tính kết nối ngang cấp cho các mạng gói. Giao thức khởi tạo phiên (SIP) đang được xem như giao thức báo hiệu chính trong hệ thống chuyển mạch mềm (softswitch). Hình 4.1 dưới đây là mô tả chồng giao thức sử dụng cho VoIP.



Hình 4.1: Chồng giao thức VoIP

Giao thức H.323 phiên bản 1 và 2 hỗ trợ H.245 trên nền TCP, Q.931 trên nền TCP và RAS trên nền UDP. Các phiên bản 3 và 4 của H.323 hỗ trợ thêm H.245 và Q.931 trên nền UDP. Giao thức SIP hỗ trợ cả TCP và UDP. Trong mạng NGN các cuộc gọi thoại đều là các cuộc gọi VoIP. Chính vì thế phần tiếp theo sẽ đề cập đến các giao thức báo hiệu trong mạng VoIP một cách chi tiết hơn.

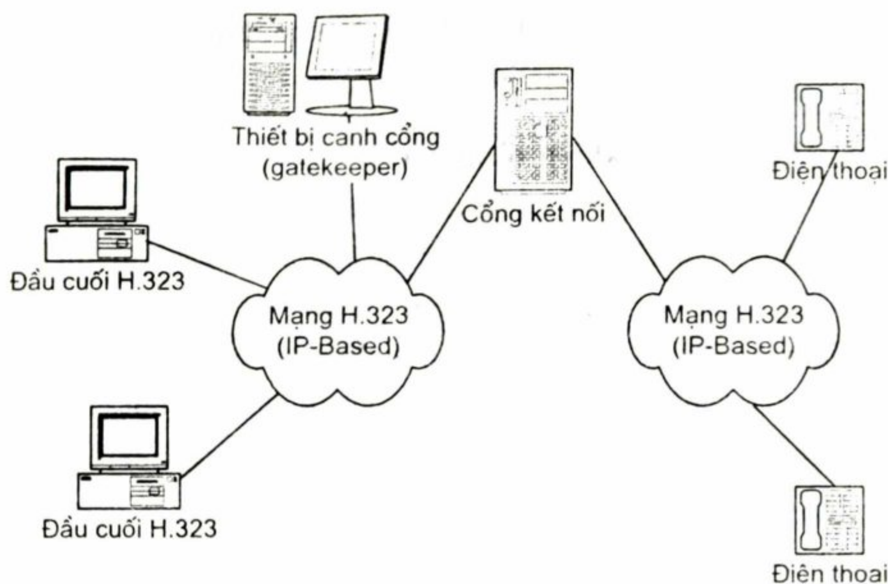
4.2. BỘ GIAO THỨC H.323

H.323 là chuẩn mở được ITU-T phát triển cho việc điều khiển cuộc gọi ngang hàng, dựa trên cơ sở của H.320 và ISDN Q.931. H.323 là một cấu trúc chặt chẽ, phức tạp và phù hợp với việc thực thi các đặc tính thoại truyền thống. Tiêu chuẩn H.323 thiết kế cho truyền audio, video và dữ liệu qua mạng IP bao gồm Internet. Tiêu chuẩn H.323 bao gồm báo hiệu và điều khiển cuộc gọi, truyền và điều khiển đa phương tiện và điều khiển băng thông cho hội nghị điểm - điểm và đa điểm. Tiêu chuẩn H.323 bao gồm các giao thức được liệt kê trong bảng 4.1.

Bảng 4.1: Bộ giao thức H.323

Đặc điểm	Giao thức
Báo hiệu cuộc gọi (Call Signalling)	H.225
Điều khiển phương tiện (Media Control)	H.245
Bộ codec âm thanh (Audio Codecs)	G.711, G.722, G.723.1, G.728, G.729
Bộ codec vi-đê-ô (Video Codecs)	H.261, H.263
Chia sẻ dữ liệu (Data Sharing)	T.120
Truyền tải phương tiện (Media Transport)	RTP/RTCP

Các thành phần cơ bản của hệ thống mạng H.323 được mô tả trong hình 4.2 dưới đây. Bao gồm các đầu cuối, cổng kết nối, thiết bị điều khiển cổng nối (gatekeeper) và khối điều khiển đa điểm (MCU).



Hình 4.2: Các thành phần mạng H.323

4.2.1. Thiết bị đầu cuối

Đầu cuối cung cấp thông tin điểm - điểm và đa điểm với các đầu cuối khác. Đầu cuối H.323 gồm các khối như điều khiển hệ thống, khối truyền tải phương tiện, mã hóa audio và giao diện với mạng IP. Phần thiết bị tùy chọn có thể là mã hóa video và thiết bị truyền dữ liệu.

4.2.2. Cổng kết nối

Cổng kết nối cung cấp giao diện giữa hai mạng khác nhau. Cổng kết nối H.323 kết nối mạng H.323 với mạng không phải H.323 như PSTN. Cổng kết nối chuyển đổi giữa audio (âm thanh), video và các định dạng truyền dữ liệu cũng như các giao thức và hệ thống thông tin. Cổng kết nối chỉ cần thiết khi

phải kết nối với mạng phi H.323, do đó không cần thiết khi kết nối giữa hai thiết bị đầu cuối H.323.

4.2.3. Thiết bị điều khiển cổng nối

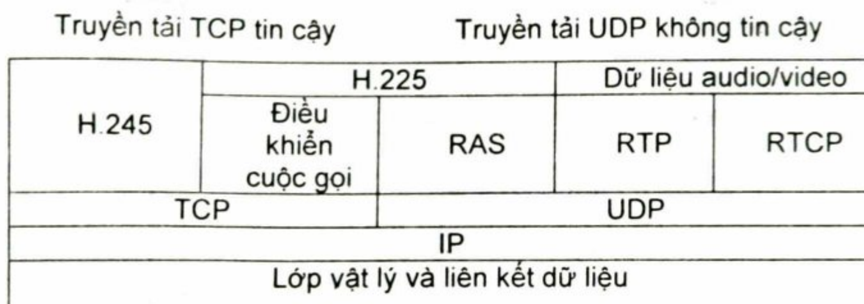
Thiết bị điều khiển cổng nối là tuỳ chọn, có thể sử dụng hoặc không. Thiết bị điều khiển cổng nối cung cấp các dịch vụ trước khi diễn ra cuộc gọi và dịch vụ điều khiển cuộc gọi cho các điểm cuối H.323. Tuy nhiên, nếu thiết bị điều khiển cổng nối có mặt trong mạng nó sẽ có nhiệm vụ: biên dịch địa chỉ, điều khiển chấp nhận, điều khiển băng thông và quản lý vùng.

Các chức năng tuỳ chọn của thiết bị điều khiển cổng nối bao gồm: báo hiệu điều khiển cuộc gọi, xác thực cuộc gọi, quản lý băng thông, quản lý cuộc gọi.

4.2.4. Khối điều khiển đa điểm (MCU)

Khối điều khiển đa điểm là điểm cuối (endpoint) hỗ trợ hội nghị giữa 3 thành viên hoặc nhiều hơn. MCU điển hình bao gồm bộ điều khiển đa điểm (MC) và một hoặc nhiều bộ xử lý đa điểm (MP). MC xử lý điều khiển và báo hiệu để hỗ trợ hội nghị trong khi MP nhận dữ liệu audio, video và luồng dữ liệu, xử lý và phân bố chúng đến các điểm cuối thành viên trong hội nghị đa điểm.

Bộ giao thức H.323 bao gồm nhiều giao thức. Bộ giao thức hỗ trợ chấp nhận cuộc gọi, thiết lập, trạng thái, giải phóng, luồng phương tiện và các bản tin trong hệ thống H.323. Các giao thức được hỗ trợ cả cơ chế truyền đưa ra gói tin cậy và không tin cậy qua mạng IP như minh họa trong hình 4.3.



Hình 4.3: Các lớp của bộ giao thức H.323

Bộ giao thức H.323 bao gồm ba vùng điều khiển:

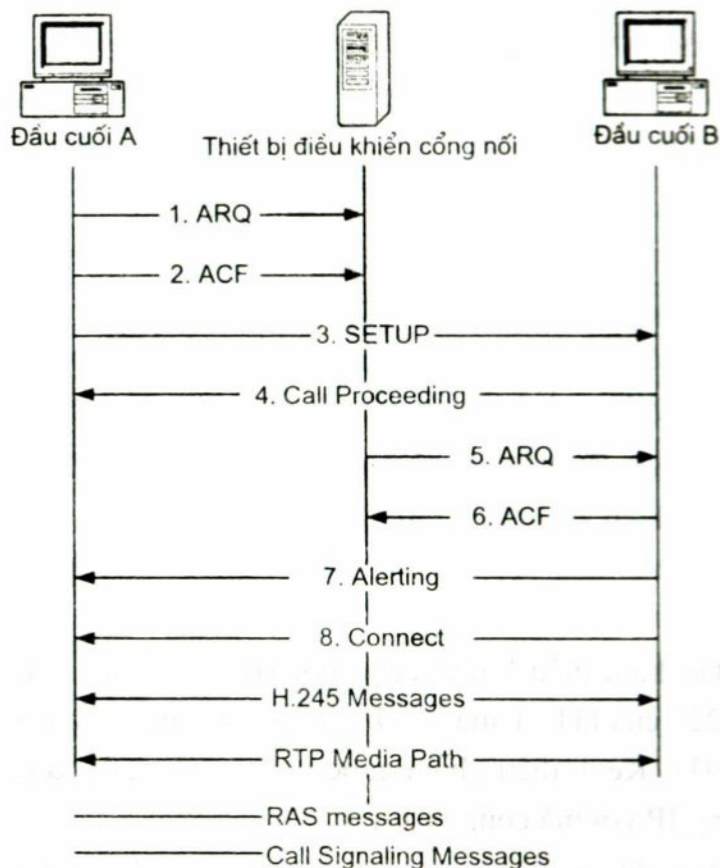
- Báo hiệu đăng ký, thừa nhận và trạng thái (RAS): Báo hiệu RAS cung cấp điều khiển trước cuộc gọi trong mạng thiết bị điều khiển cổng nối H.323.

- Báo hiệu điều khiển cuộc gọi (H.225): dựa trên khuyến nghị H.225 của ITU-T mà nó chỉ ra việc sử dụng các bản tin bản hiệu Q.931. Kênh điều khiển cuộc gọi tin cậy TCP được tạo ra trên mạng IP với mã cổng là 1720.

- Điều khiển và truyền tải thông tin Media (H.245 và RTP/RTCP): H.245 xử lý các bản tin điều khiển từ đầu đến cuối giữa các điểm cuối H.323. Thủ tục giao thức H.245 thiết lập kênh cho truyền audio, video dữ liệu và thông tin kênh điều khiển. Truyền đa phương tiện trong H.323 được cung cấp bởi RTP và RTCP.

4.2.5. Trình tự thiết lập cuộc gọi

Nếu xem xét một cách chi tiết thì cuộc gọi giữa hai đầu cuối H.323 được thiết lập như hình 4.4 dưới đây.



Hình 4.4: Thiết lập cuộc gọi H.323

- Trước hết cả 2 phải đã được đăng ký tại thiết bị điều khiển cổng nối

- Đầu cuối A gửi yêu cầu tới thiết bị điều khiển cổng nối đề nghị thiết lập cuộc gọi

- Thiết bị điều khiển cổng nối gửi cho đầu cuối A thông tin cần thiết về đầu cuối B

- Đầu cuối A gửi bản tin SETUP tới đầu cuối B

- Đầu cuối B trả lời bằng bản tin Call Proceeding và đồng thời liên lạc với thiết bị điều khiển cổng nối để xác nhận quyền thiết lập cuộc gọi

- Đầu cuối B gửi bản tin cảnh báo và kết nối (Alerting and Connect)

- Hai đầu cuối trao đổi một số bản tin H.245 để xác định chủ/tớ, khả năng xử lý của đầu cuối và thiết lập kế nối RTP.

Đây là trường hợp cuộc gọi điểm-điểm đơn giản nhất, khi mà báo hiệu cuộc gọi không được định tuyến tới thiết bị điều khiển cổng nối. H.323 hỗ trợ nhiều kịch bản thiết lập cuộc gọi phức tạp khác.

H.323 là hệ thống ghép lai được xây dựng từ các thiết bị tập trung thông minh như: thiết bị điều khiển cổng nối, MCU, cổng kết nối và điểm cuối. Mặc dù chuẩn H.323 trong phiên bản gần đây nhất có phần hoàn thiện hơn song vấn đề vẫn nảy sinh, như thời gian thiết lập cuộc gọi dài, quá nhiều chức năng thiết bị điều khiển cổng nối phải thực hiện và khả năng mở rộng khi sử dụng kiểu báo hiệu cuộc gọi định tuyến qua thiết bị điều khiển cổng nối (GKRCS).

Khi cần sử dụng cổng kết nối dung lượng lớn để kết nối mạng PSTN, người ta sẽ sử dụng giao thức điều khiển cổng đơn giản (SGCP: Simple Gateway Control Protocol) và Giao thức điều khiển cổng phương tiện (MGCP: Media Gateway Control Protocol) để thay thế giao thức cho cổng kết nối H.323. Các hệ thống điều khiển cuộc gọi này có vẻ hiệu quả hơn, đáp ứng nhu cầu của các nhà cung cấp cỡ lớn.

Cũng vậy, giao thức khởi tạo phiên SIP (Session Initiation Protocol) có thể giải quyết một số tồn tại của H.323 và nó khả năng sẽ thay thế H.323.

4.3. GIAO THỨC KHỞI TẠO PHIÊN (SIP)

SIP là giao thức điều khiển báo hiệu thuộc lớp ứng dụng, được phát triển như là một chuẩn mở RFC 2543 của IETF. Khác với H.323, nó dựa trên nguồn gốc Web (HTTP) và có thiết kế kiểu mô-đun, đơn giản và dễ dàng mở rộng với các ứng dụng thoại SIP. SIP là một giao thức báo hiệu để thiết lập, duy trì và kết thúc các phiên đa phương tiện như: thoại IP, hội nghị và các ứng dụng tương tự khác liên quan đến việc truyền thông tin đa phương tiện.

Ưu điểm chính của SIP so với các giao thức báo hiệu khác là cung cấp một sự mềm dẻo. Nó được thiết kế nhanh và đơn giản. Giao thức SIP cung cấp các chức năng sau:

- Định vị người dùng thông qua địa chỉ tương tự như email.

- Năng lực người dùng: các tham số phiên có thể thương lượng giữa hai phía.

- Lợi ích người dùng: Xác định dựa trên kiểu bên bị gọi muốn tiến hành truyền thông.

Giao thức SIP bao gồm hai thành phần: đại lý trạm của người sử dụng (user agent) và máy chủ mạng (network server). User agent là một ứng dụng kết cuối hệ thống mà nó bao gồm cả user agent client (UAC) khởi tạo cuộc gọi và user agent server (UAS) nó sẽ trả lời cuộc gọi. Kiến trúc của SIP cho phép thông tin ngang cấp sử dụng giao thức máy khách/máy chủ.

Network servers bao gồm bốn kiểu là: Proxy server (máy chủ ủy quyền), Location server (máy chủ cục bộ), Redirect

server (máy chủ chuyển tiếp) và Register server (máy chủ đăng ký).

ITU-T không phải là tổ chức tiêu chuẩn duy nhất đưa ra kế hoạch thiết lập kết nối thoại IP và đóng gói audio. Hơn hẳn bất kỳ một tổ chức quản lý các chuẩn mực Internet, IETF cũng có những yêu cầu của riêng mình đối với những hệ thống VoIP và được gọi là "Giao thức điều khiển phiên" - SIP.

Những người đề xuất SIP cho rằng H.323 đang xuất hiện trong báo hiệu ATM và ISDN, là không thích hợp cho điều khiển hệ thống VoIP nói chung và trong thoại Internet nói riêng. Yêu cầu này khẳng định là H.323 vốn dĩ rất phức tạp, hỗ trợ các chức năng phần lớn là không cần thiết cho thoại IP do đó đòi hỏi chi phí cao và không hiệu quả. Ví dụ H.323 xác định 3 phương pháp khác nhau để phối hợp hoạt động giữa H.225 và H.245, với các kết nối khác nhau, H.245 ngang qua kết nối H.225 và tiến hành phương pháp "kết nối nhanh" của 2 giao thức tích hợp. Mặc dù hầu hết các khả năng thực hiện chỉ hỗ trợ cho các kết nối nhanh, tính tương thích H.323 liên quan đến yêu cầu hỗ trợ của cả 3 phương pháp.

Đồng thời, những người ủng hộ SIP cũng cho rằng H.323 không có khả năng mở rộng yêu cầu đối với giao thức báo hiệu cho công nghệ chẳng hạn như VoIP, là những công nghệ chắc chắn sẽ phát triển và hỗ trợ các dịch vụ và đặc tính mới. Kinh nghiệm trong sử dụng các giao thức Internet mail (SMTP) đã cung cấp rất nhiều triết lý phát triển SIP, trong đó tập trung vào khả năng thích ứng của báo hiệu trong tương lai.

SIP dựa trên ý tưởng và cấu trúc của HTTP (Hypertext Transfer Protocol), giao thức trao đổi thông tin của World Wide Web. Nó được định nghĩa như một giao thức Client-Server, trong đó các yêu cầu được bên gọi (client) đưa ra và bên bị gọi (server) trả lời. SIP sử dụng một số kiểu bản tin và các trường mào đầu của HTTP, xác định nội dung luồng thông tin theo mào đầu thực thể (mô tả nội dung - kiểu loại) và cho phép xác nhận các phương pháp sử dụng giống nhau được sử dụng trên Web.

SIP định nghĩa các bản tin *INVITE* và *ACK* giống như bản tin *Setup* và *Connect* trong H.225, trong đó cả hai đều định nghĩa quá trình mở 1 kênh đáng tin cậy mà thông qua đó cuộc gọi có thể đi qua. Tuy nhiên khác với H.225, độ tin cậy của kênh này không phụ thuộc vào TCP. Các nhà đề xuất SIP kiến nghị rằng việc tích hợp độ tin cậy vào lớp ứng dụng này cho phép kết hợp một cách chặt chẽ các giá trị điều chỉnh để ứng dụng, có thể tối ưu hóa VoIP hơn là phụ thuộc vào những giá trị "mục đích chung chung" của TCP.

Cuối cùng, SIP dựa vào giao thức mô tả phiên (SDP), một tiêu chuẩn khác của IETF, để thực hiện sự sắp xếp tương tự theo cơ cấu chuyển đổi dung lượng của H.245. SDP được dùng để nhận dạng mã tổng đài trong những cuộc gọi sử dụng một mô tả nguyên bản đơn. SDP cũng được sử dụng để chuyển các phần tử thông tin của giao thức báo hiệu thời gian thực RTSP để sắp xếp các tham số hội nghị đa điểm và định nghĩa khuôn dạng chung cho nhiều loại thông tin khi được chuyển trong SIP.

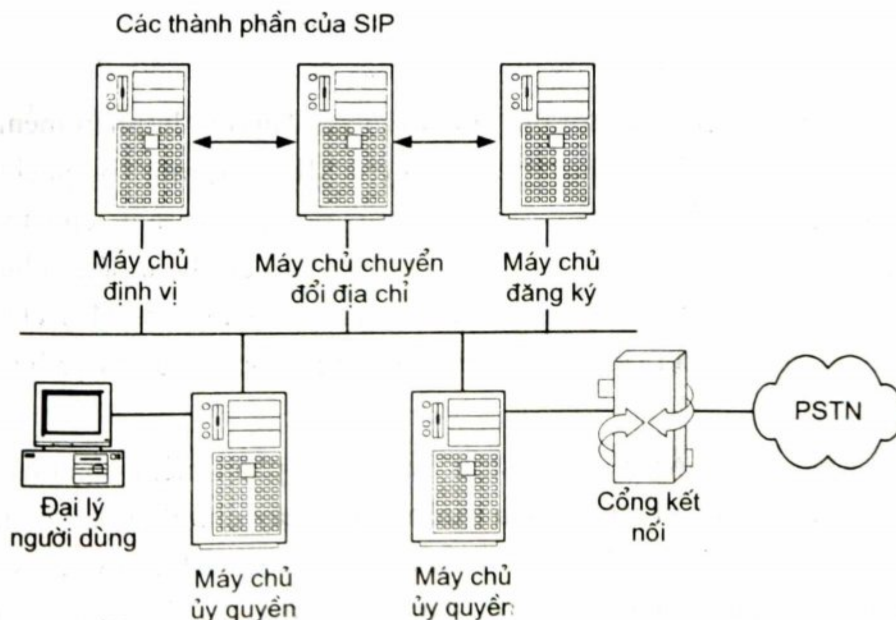
Cuộc tranh luận liệu chuẩn H.323/H.225 hay SIP phù hợp cho thoại Internet hay không thì chỉ mang tính chất học thuật. Cả hai cách đều có những thuận lợi và bất lợi. Trong khi H.323 có thuận lợi rất lớn trong thị trường (đã xuất hiện trước SIP) thì SIP

có lợi thế của IETF là một trong số các tổ chức tiêu chuẩn hoạt động tích cực. Với mục đích nghiên cứu các giao thức và giải pháp cho vấn đề VoIP, tốt nhất là lưu ý đến cả H.323 và SIP.

Theo định nghĩa của IETF, SIP là "giao thức báo hiệu lớp ứng dụng mô tả việc khởi tạo, thay đổi và huỷ các phiên kết nối tương tác đa phương tiện giữa những người sử dụng". SIP có thể sử dụng cho rất nhiều dịch vụ khác nhau trong mạng IP như dịch vụ thoại, hội nghị thoại, email, dạy học từ xa, quảng bá (MPEG, MP3...), truy nhập HTML, XML, hội nghị truyền hình...

4.3.1. Phần mềm chuyển mạch cuộc gọi SIP

Phần mềm chuyển cuộc gọi SIP có cấu trúc như sau:



Hình 4.5: Cấu trúc SIP

Trong hình trên, đại lý người dùng (*User Agent)* là thiết bị đầu cuối trong mạng SIP, có thể là một máy điện thoại SIP, có thể là máy tính chạy phần mềm đầu cuối SIP.

Máy chủ ủy quyền *(Proxy Server)* là phần mềm trung gian hoạt động vừa như máy chủ và vừa như máy trạm để thực hiện các yêu cầu thay mặt các đầu cuối khác. Tất cả các yêu cầu được xử lý tại chỗ bởi máy chủ ủy quyền nếu có thể, hoặc được chuyển cho các máy chủ khác. Trong trường hợp máy chủ ủy quyền không trực tiếp đáp ứng các yêu cầu này thì máy chủ ủy quyền sẽ thực hiện khâu chuyển đổi hoặc dịch sang khuôn dạng thích hợp trước khi chuyển đi.

Máy chủ định vị *(Location Server)* là phần mềm định vị thuê bao, cung cấp thông tin về những vị trí có thể của phía bị gọi cho các phần mềm máy chủ ủy quyền và máy chủ chuyển đổi địa chỉ.

Máy chủ chuyển đổi địa chỉ *(Redirect Server)* là phần mềm nhận yêu cầu SIP và chuyển đổi địa chỉ SIP sang một số địa chỉ khác và gửi lại cho đầu cuối. Không giống như máy chủ ủy quyền, máy chủ chuyển đổi địa chỉ không bao giờ hoạt động như một đầu cuối, tức là không gửi đi bất cứ yêu cầu nào. Máy chủ chuyển đổi địa chỉ (Redirect Server) cũng không nhận hoặc huỷ cuộc gọi.

Máy chủ đăng ký (Registrar Server) là phần mềm nhận các yêu cầu đăng ký. Trong nhiều trường hợp máy chủ đăng ký đảm nhiệm luôn một số chức năng an ninh như xác nhận người sử dụng. Thông thường máy chủ đăng ký được cài đặt cùng với máy chủ ủy quyền và máy chủ hay địa chỉ hoặc cung cấp dịch vụ định vị thuê bao. Mỗi lần đầu cuối được bật lên (thí dụ máy

điện thoại hoặc phần mềm SIP) thì đầu cuối lại đăng ký với máy chủ. Nếu đầu cuối cần thông báo cho máy chủ về địa điểm của mình thì bản tin REGISTER cũng được gửi đi. Nói chung các đầu cuối đều thực hiện việc đăng ký lại một cách định kỳ.

4.3.2. Các bản tin SIP, mào đầu và đánh số

Dưới đây là các bản tin của SIP:

- INVITE: Bắt đầu thiết lập cuộc gọi bằng cách gửi bản tin mời đầu cuối khác tham gia.

- ACK: Bản tin này khẳng định máy trạm đã nhận được bản tin trả lời bản tin INVITE.

- BYE: Bắt đầu kết thúc cuộc gọi

- CANCEL: Hủy yêu cầu đang nằm trong hàng đợi

- REGISTER: Đầu cuối SIP sử dụng bản tin này để đăng ký với máy chủ đăng ký.

- OPTIONS: Sử dụng để xác định năng lực của máy chủ.

- INFO: Sử dụng để tải các thông tin như âm báo DTMF.

Giao thức SIP có nhiều điểm trùng hợp với giao thức HTTP. Các bản tin trả lời các bản tin SIP nêu trên gồm có:

- 1xx – Các bản tin chung

- 2xx – Thành công

- 3xx – Chuyển địa chỉ

- 4xx – Yêu cầu không được đáp ứng

- 5xx – Sự cố của máy chủ

- 6xx – Sự cố toàn mạng.

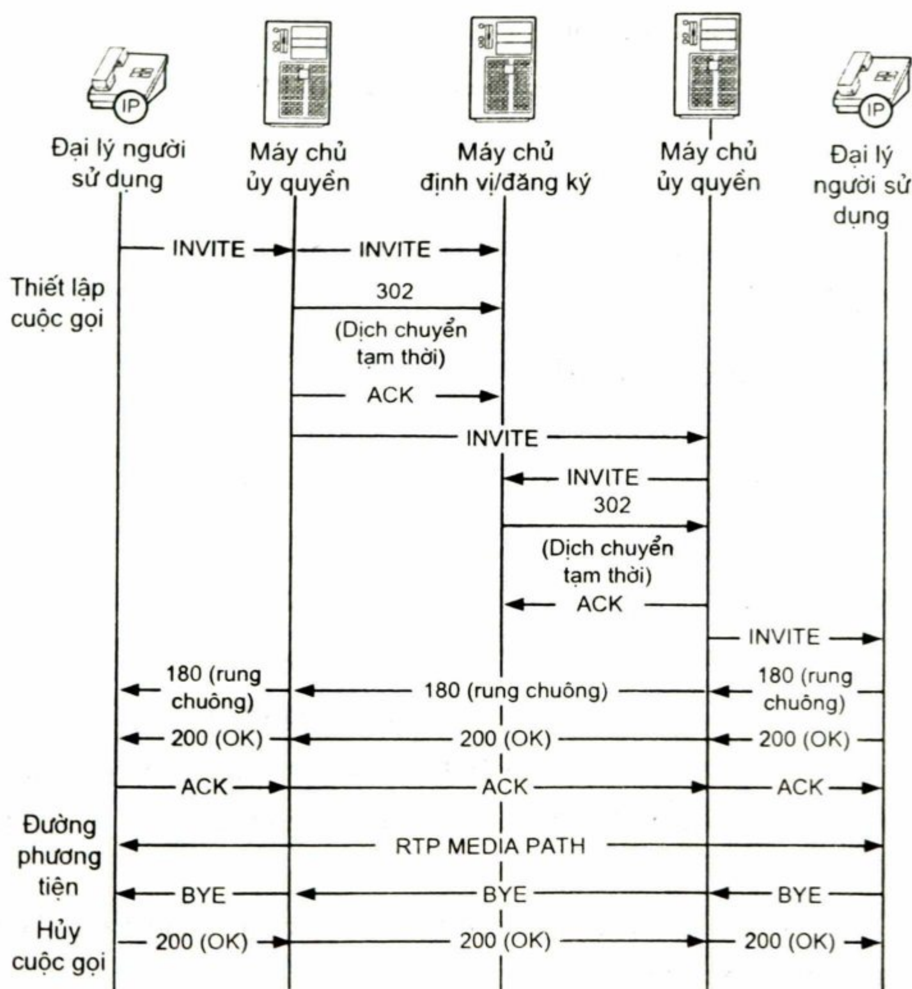
Các bản tin SIP đều có khuôn dạng text, tương tự như HTTP. Mào đầu của bản tin SIP cũng tương tự như HTTP và SIP cũng hỗ trợ MIME (một số chuẩn về email). Sau đây là thí dụ về mào đầu của bản tin SIP:

```
------------------------------------------------------------
SIP Header
------------------------------------------------------------
INVITE sip:5120@192.168.36.180 SIP/2.0
Via: SIP/2.0/UDP 192.168.6.21:5060
From: sip:5121@192.168.6.21
To: <sip:5120@192.168.36.180>
Call-ID: c2943000-e0563-2a1ce-2e323931@192.168.6.21
CSeq: 100 INVITE
Expires: 180
User-Agent: Cisco IP Phone/ Rev. 1/ SIP enabled
Accept: application/sdp
Contact: sip:5121@192.168.6.21:5060
Content-Type: application/sdp
```

Các đầu cuối SIP được đánh số tương tự như địa chỉ email, thí dụ như ptit@vnpt.com.vn.

4.3.3. Thiết lập và huỷ cuộc gọi SIP

Trong mạng SIP quá trình thiết lập và hủy một phiên kết nối thường gồm có 6 bước như sau:



Hình 4.6: Thiết lập và huỷ cuộc gọi SIP

1. Đăng ký, khởi tạo và định vị đầu cuối

2. Xác định phương tiện của cuộc gọi, tức là mô tả phiên mà đầu cuối được mời tham dự

3. Xác định mong muốn của đầu cuối bị gọi, trả lời hay không. Phía bị gọi phải gửi bản tin xác nhận chấp thuận cuộc gọi hoặc từ chối

4. Thiết lập cuộc gọi

5. Thay đổi hay điều khiển cuộc gọi (thí dụ như chuyển cuộc gọi)

6. Hủy cuộc gọi.

Quá trình thiết lập và huỷ cuộc gọi SIP cơ bản được mô tả trong hình 4.6 sau.

4.3.4. Tính năng của SIP

Giao thức SIP được thiết kế với những tiêu chí sau:

- Tích hợp với các giao thức đã có của IETF
- Đơn giản và có khả năng mở rộng
- Hỗ trợ tối đa sự di động của đầu cuối
- Dễ dàng tạo tính năng mới cho dịch vụ.

a) Tích hợp với các giao thức đã có của IETF

Các giao thức khác của IETF có thể sử dụng để xây dựng những ứng dụng SIP. SIP có thể hoạt động cùng với nhiều giao thức như:

- RSVP (Resource Reservation Protocol): Giao thức dành trước tài nguyên mạng

- RTP (Real-time transport Protocol): Giao thức truyền tải thời gian thực

- RTSP (Real Time Streaming Protocol): Giao thức tạo luồng thời gian thực

- SAP (Session Advertisement Protocol): Giao thức thông báo trong phiên kết nối

- SDP (Session Description Protocol): Giao thức mô tả phiên kết nối đa phương tiện

- MIME (Multipurpose Internet Mail Extension - Mở rộng thư tín Internet đa mục đích): Giao thức thư điện tử

- HTTP (Hypertext Transfer Protocol): Giao thức truyền siêu văn ban

- COPS (Common Open Policy Service): Dịch vụ chính sách mở chung.

- OSP (Open Settlement Protocol): Giao thức thỏa thuận mở

b) Đơn giản và có khả năng mở rộng

SIP có rất ít bản tin, không có các chức năng thừa nhưng SIP có thể sử dụng để thiết lập những phiên kết nối phức tạp như hội nghị... Đơn giản, gọn nhẹ, dựa trên khuôn dạng văn bản. SIP là giao thức ra đời sau và đã khắc phục được điểm yếu của nhiều giao thức trước đây.

Các phần mềm của máy chủ ủy quyền, máy chủ đăng ký, máy chủ chuyển đổi địa chỉ, máy chủ định vị... có thể chạy trên các máy chủ khác nhau và việc cài đặt thêm máy chủ hoàn toàn không ảnh hưởng đến các máy chủ đã có. Chính vì thế hệ thống chuyển mạch SIP có thể dễ dàng nâng cấp.

c) Hỗ trợ tối đa sự di động của đầu cuối

Do có máy chủ ủy quyền, máy chủ đăng ký và máy chủ chuyển đổi địa chỉ hệ thống luôn nắm được địa điểm chính xác của thuê bao. Thí dụ thuê bao với địa chỉ ptit@vnpt.com.vn có thể

nhận được cuộc gọi thoại hay thông điệp ở bất cứ địa điểm nào qua bất cứ đầu cuối nào như máy tính để bàn, máy xách tay, điện thoại SIP... Với SIP rất nhiều dịch vụ di động mới được hỗ trợ.

d) Dễ dàng tạo tính năng mới cho dịch vụ và dịch vụ mới

Là giao thức khởi tạo phiên trong mạng chuyển mạch gói SIP cho phép tạo ra những tính năng mới hay dịch vụ mới một cách nhanh chóng. Ngôn ngữ xử lý cuộc gọi (Call Processing Language) và Giao diện cổng kết nối chung (Common Gateway Interface) là một số công cụ để thực hiện điều này. SIP hỗ trợ các dịch vụ thoại như chờ cuộc gọi, chuyển tiếp cuộc gọi, khóa cuộc gọi... (call waiting, call forwarding, call blocking...), hỗ trợ thông điệp thống nhất...

4.3.5. So sánh H.323 và SIP

Giữa H.323 và SIP có nhiều điểm tương đồng. Cả hai đều cho phép điều khiển, thiết lập và huỷ cuộc gọi. Cả H.323 và SIP đều hỗ trợ tất cả các dịch vụ cần thiết, tuy nhiên có một số điểm khác biệt giữa hai chuẩn này. Đó là:

- H.323 hỗ trợ hội nghị đa phương tiện rất phức tạp. Hội nghị H.323 về nguyên tắc có thể cho phép các thành viên sử dụng những dịch vụ như bảng thông báo, trao đổi dữ liệu, hoặc hội nghị video.

- SIP hỗ trợ SIP-CGI (SIP - Common Gateway Interface).

- SIP hỗ trợ điều khiển cuộc gọi từ một đầu cuối thứ 3. Hiện nay H.323 đang được nâng cấp để hỗ trợ chức năng này.

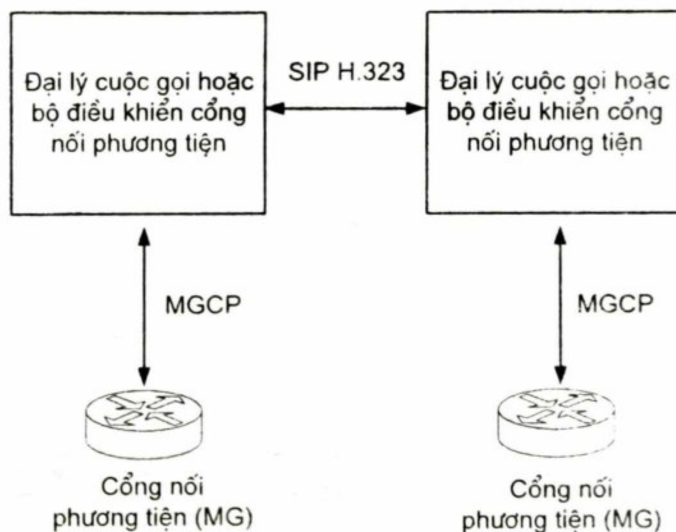
Bảng 4.2: So sánh H.323 và SIP

	SIP	**H.323**
Tổ chức	IETF	ITU
Quan hệ kết nối	Ngang cấp	Ngang cấp
Khởi điểm	Dựa trên mạng Internet và Web. Cú pháp và bản tin tương tự như HTTP.	Cơ sở là mạng thoại. Giao thức báo hiệu tuân theo chuẩn ISDN Q.SIG.
Đầu cuối	Đầu cuối thông minh SIP	Đầu cuối thông minh H.323
Các máy chủ lõi	Máy chủ ủy quyền SIP, máy chủ chuyển đổi địa chỉ, máy chủ định vị và máy chủ đăng ký.	Điều khiển cổng nối H.323
Tình hình hiện nay	Giai đoạn thử nghiệm khả năng cùng hoạt động của thiết bị của các nhà cung cấp khác nhau đã kết thúc. SIP nhanh chóng trở nên phổ biến	Đã được sử dụng rộng rãi.
Khuôn dạng bản tin	Text, UTF-8	Nhị phân ASN.1 PER
Trễ thiết lập cuộc gọi	1,5 RTT (round-trip time, tức chu kỳ gửi bản tin và nhận bản tin trả lời hay xác nhận)	6-7 RTT hoặc hơn
Giám sát trạng thái cuộc gọi	Có 2 lựa chọn: chỉ trong thời gian thiết lập cuộc gọi hoặc suốt thời gian cuộc gọi	Phiên bản 1 và 2: máy chủ phải giám sát trong suốt thời gian cuộc gọi và phải giữ trạng thái kết nối TCP -> hạn chế khả năng mở rộng và giảm độ tin cậy
Báo hiệu quảng bá (cho cả nhóm...)	Có hỗ trợ	Không

Chất lượng dịch vụ	Sử dụng các giao thức khác như RSVP, OPS, OSP để đảm bảo chất lượng dịch vụ	Thiết bị điều khiển cổng nối điều khiển băng thông. H.323 khuyến nghị dùng RSVP để lưu trữ tài nguyên mạng.
Bảo mật	Đăng ký tại máy chủ đăng ký, có xác nhận đầu cuối và mã hóa	Chỉ đăng ký khi trong mạng có thiết bị điều khiển cổng nối, xác nhận và mã hóa theo chuẩn H.235.
Định vị đầu cuối và định tuyến cuộc gọi	Dùng SIP URL để đánh địa chỉ. Định tuyến nhờ sử dụng máy chủ thay địa chỉ và máy chủ định vị.	Định vị đầu cuối sử dụng E.164 hoặc tên ảo H.323 và phương pháp ánh xạ địa chỉ nếu trong mạng có thiết bị điều khiển cổng nối. Chức năng định tuyến do thiết bị điều khiển cổng nối đảm nhiệm.
Tính năng thoại	Hỗ trợ các tính năng của cuộc gọi cơ bản	Hỗ trợ các tính năng của cuộc gọi cơ bản
Hội nghị	Hội nghị cơ sở, quản lý phân tán	Được thiết kế nhằm hỗ trợ rất nhiều tính năng hội nghị, kể cả thoại, video và dữ liệu, quản lý tập trung nên MC có thể tắc nghẽn
Tạo tính năng và dịch vụ mới	Dễ dàng, sử dụng SIP-CGI và CPL	H.450.1
Khả năng mở rộng	Dễ dàng	Hạn chế
Tích hợp với web	Rất tốt, hỗ trợ nhấn chuột để quay số (click-to-dial)	Kém

4.4. GIAO THỨC MGCP

Giao thức điều khiển cổng nối phương tiện (MGCP) là giao thức dùng để điều khiển các cổng nối thoại từ các thiết bị điều khiển cuộc gọi. Giao thức này là chức năng của bộ điều khiển cổng nối phương tiện (MGC) hoặc đại lý cuộc gọi (CA).

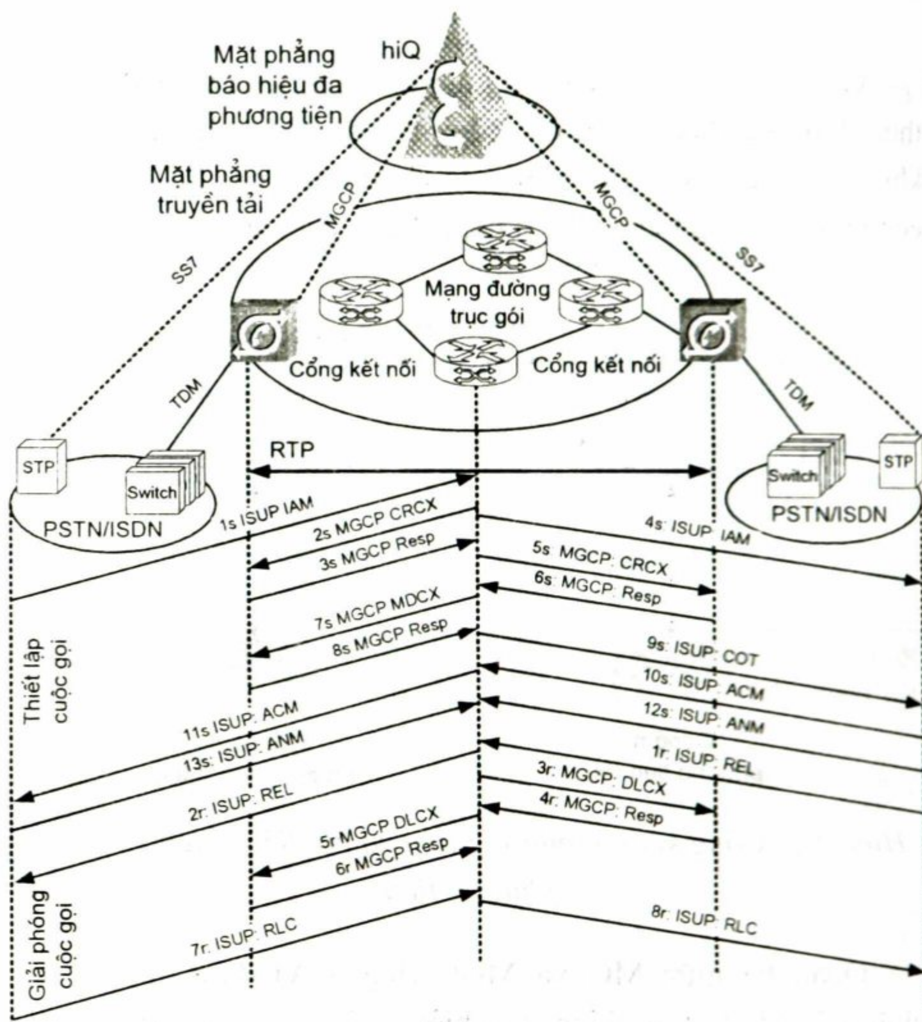


Hình 4.7: Cổng kết nối phương tiện và bộ điều khiển cổng nối phương tiện

Quan hệ giữa MG và MGC (hay CA) được mô tả trên hình 4.7. MGC thực hiện báo hiệu cuộc gọi, điều khiển MG. MGC và MG trao đổi lệnh với nhau thông qua MGCP.

4.4.1. Thiết lập cuộc gọi

Hình 4.8 mô tả việc thiết lập và giải phóng cuộc gọi MGCP thông thường giữa 2 thuê bao PSTN kết nối với các tổng đài sử dụng CCS7 để báo hiệu.



Hình 4.8: Thiết lập và giải phóng cuộc gọi MGCP

- Thuê bao PSTN bên A (chủ gọi) nhấc máy, quay số rồi gửi các chữ số địa chỉ thuê bao bị gọi này tới tổng đài (bên chủ gọi – bên A).

- Tổng đài bên A sau khi nhận xong địa chỉ thuê bao bị gọi (bên B), tạo ra bản tin ISUP IAM (Initial Address Message) để

gửi báo hiệu cho cổng nối phương tiện/thiết bị điều khiển cổng nối phương tiện (MG/MGC) để chính thức đặt vấn đề kết nối với thuê bao bị gọi (1)

- MGC gửi cho MG bên A bản tin MGCP **CRCX - "Createconnection"** (2)

- MG bên A trả lời bằng bản **Resp - "Response"** (3)

- MGC gửi cho tổng đài bên B (bị gọi) bản tin ISUP **IAM** (4)

- MGC gửi tiếp bản tin MGCP **CRCX** cho MG bên B (5)

- MG bên B trả lời bằng bản tin **Resp (6)**

- MGC gửi cho MG bên A bản tin MGCP **MDCX** (7)

- MG bên A trả lời bằng bản **Resp (8)**

- MGC gửi cho MG bên B bản tin ISUP **COT** (Continuity message) để báo cáo kết quả kiểm tra tính liên tục (được yêu cầu trong bản tin IAM) trên một mạch (kênh B được chọn) (9)

- Tổng đài bên B gửi cho MGC bản tin ISUP **ACM** (hoàn tất việc nhận địa chỉ) đồng thời rung chuông cho thuê bao B (10)

- MGC chuyển cho tổng đài bên A bản tin ISUP **ACM** này (11)

- Sau khi thuê bao B nhấc máy, tổng đài bên B gửi cho MGC bản tin ISUP **ANM** (ANswer Message) thông báo việc nhấc máy này (12)

- Bản tin **ANM** này lại được MGC chuyển cho tổng đài bên A và hai bên thuê bao A và B có thể bắt đầu đàm thoại (13).

Dưới đây là thủ tục giải phóng kết nối MGCP giữa 2 thuê bao A và B khi một trong hai bên giả sử bên B đặt máy

- Tổng đài bên B gửi bản tin ISUP **REL** cho MGC yêu cầu giải phóng kết nối (1)

- MGC sau đó gửi bản tin ISUP **REL** này cho tổng đài bên A (2)

- MGC gửi bản tin MGCP **DLCX "Deleteconnection"**cho MG bên B để thông báo yêu cầu MG giải phóng kết nối (3)

- MG bên B gửi trả lại bản tin MGCP **Resp (4)**

- MGC cũng đồng thời gửi bản tin MGCP **DLCX** cho MG bên A để thông báo yêu cầu MG giải phóng kết nối (5)

- MG bên A cũng gửi trả lại bản tin MGCP **Resp** (6)

- Tổng đài bên A gửi cho MGC bản tin ISUP **RLC** thông báo hoàn tất việc giải phóng kết nối (7)

- MGC sau đó cũng gửi cho tổng đài bên B bản tin ISUP **RLC** thông báo hoàn tất việc giải phóng kết nối. Kết nối kết thúc (8).

4.4.2. Các lệnh MGCP

Giao thức MGCP có 2 nhóm lệnh như sau:

a) Lệnh MGC

- EndpointConfiguration
- NotificationRequest
- CreateConnection
- ModifyConnection
- DeleteConnection
- AuditEndpoint
- AuditConnection
- DeleteConnection
- RestartInProgress

b) Lệnh cổng nối

- DeleteConnection
- RestartInProgress

4.4.3. So sánh giữa MGCP, SIP và H.323

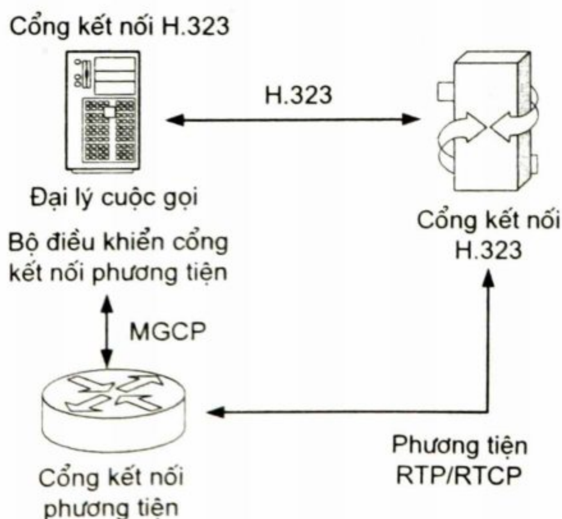
MGCP khác với SIP và H.323 ở một số điểm như sau:

- MGCP là giao thức kiểu chủ/tớ, trong khi SIP và H.323 là ngang hàng (peer-to-peer).

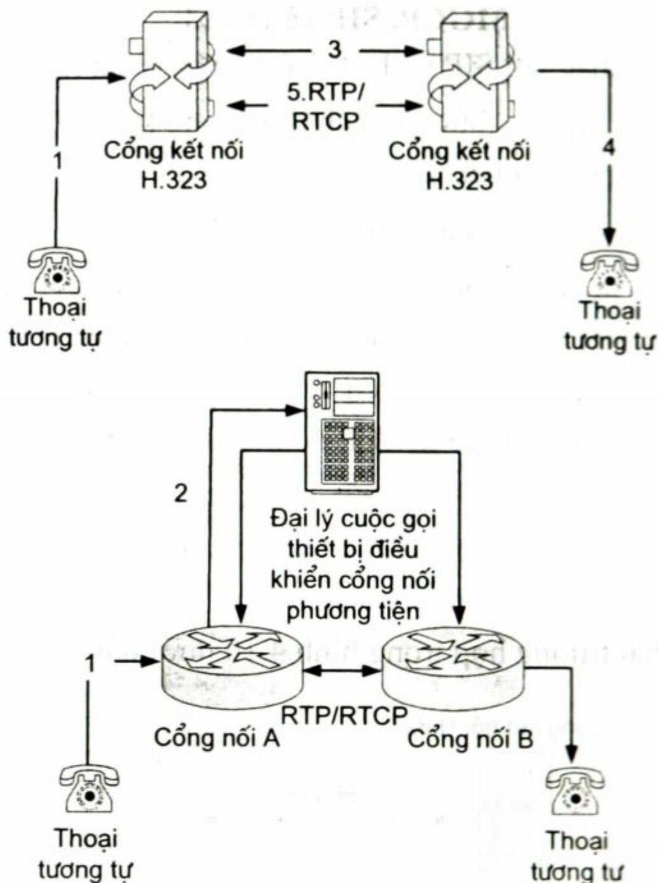
- MGCP được sử dụng giữa MG và MGC.

MGCP ra đời nhằm tách biệt các chức năng báo hiệu và chức năng thiết lập đường truyền. MGC (hay CA, chuyển mạch mềm...) sau khi nhận được yêu cầu thiết lập cuộc gọi SIP hoặc H.323 sẽ dùng giao thức MGCP để điều khiển cổng kết nối thiết lập phiên kết nối giữa 2 đầu cuối.

Trong hình 4.9 ta thấy chức năng báo hiệu đã được tách biệt và do MGC đảm nhiệm. Chúng ta xem xét việc thiết lập cuộc gọi trong hai trường hợp trong hình 4.10 dưới đây.



Hình 4.9: Cổng kết nối H.323 và MGC+MG



Hình 4.10: Báo hiệu thiết lập cuộc gọi trong hai mạng H.323 và MGCP

Trên cơ sở MGCP một số giao thức mới được phát triển, đó là

- MEGACO do IETF phát triển

- H.248 do ITU phát triển.

Các giao thức này đều là giao thức điều khiển cổng nối phương tiện (Media Gateway).

Bảng 4.3: So sánh H.323 và MGCP

H.323	MGCP
1. Thuê bao nhấc máy và quay số 2. Cổng kết nối phân tích định tuyến cuộc gọi 3. Hai cổng kết nối trao đổi thông tin 4. Cổng kết nối bị gọi đổ chuông ở số máy bị gọi 5. Hai cổng kết nối thiết lập phiên kết nối RTP/RTCP	1. Thuê bao nhấc máy và quay số 2. Cổng kết nối thông báo cho MGC 3. MGC phân tích số, định tuyến và gửi lệnh cho cổng kết nối bị gọi để đổ chuông ở số máy bị gọi 4. MGC gửi lệnh cho 2 cổng kết nối để thiết lập phiên kết nối RTP/RTCP

4.5. GIAO THỨC MEGACO/H.248

Giao thức MEGACO/H248 dựa trên mô hình chủ/tớ và là chuẩn quốc tế cho việc điều khiển cổng kết nối trong mạng phân tán và là chuẩn mở được phát triển kết hợp giữa IETF và ITU. MEGACO/H.248 tuy còn đơn giản nhưng hiệu quả và rất linh hoạt trong việc mở rộng, cho phép xây dựng phân chia các chức năng cổng kết nối bên dưới lớp điều khiển cuộc gọi (như SIP, H.323…). Nó rất linh hoạt cho việc phát triển phần lớn các dịch vụ với các yêu cầu chất lượng, giá cả khác nhau cũng như hỗ trợ và phát triển các mạng sẵn có.

Giao thức MEGACO phát triển từ giao thức điều khiển cổng đa phương tiện (MGCP- Media Gateway Control Protocol) và được định nghĩa bởi RFC 3015. Giao thức MEGACO tách điều khiển báo hiệu từ cổng nối phương tiện (MG) giữa hai mạng khác nhau và đưa chúng vào bộ điều khiển cổng nối phương tiện (MGC). MEGACO/H.248 chỉ ra giao thức thông tin giữa MGC và MG và không địa chỉ thông tin giữa các kết cuối.

Trong mạng thế hệ sau NGN, một trong các giao thức cũng có thể được sử dụng để điều khiển việc thực hiện các cuộc nối thoại đó là MEGACO/H.248. Người ta cũng gọi nó là giao thức điều khiển cổng nối phương tiện tương tự như MGCP. Giao thức này được IETF phát triển và hoàn thiện, hiện tại SG16 của ITU đã chấp nhận giao thức này dưới tên H.248.

4.5.1. Chức năng của MEGACO

Theo ITU-T SG16 và IETF WG, MEGACO là giao thức điều khiển giữa thiết bị điều khiển cổng nối phương tiện MGC và cổng nối phương tiện MG. Giao thức này có chức năng:

- Điều khiển các dạng thiết bị kết cuối.
- Hỗ trợ khả năng dàn xếp cuộc gọi
- Các phương án cuộc gọi đa người sử dụng
- Chất lượng dịch vụ và hỗ trợ cho đo lưu lượng.
- Báo lỗi trong giao thức, cuộc gọi, dung lượng và lỗi mạng

Giao thức MEGACO có thể được thiết kế theo cả dạng ngang hàng hay chủ/tớ, các giao thức này có thể cùng tồn tại trong một mạng. MEGACO hoàn toàn tương thích với các cách tiếp cận theo kiểu ngang hàng như SIP, H323, và nó cũng có những ưu điểm trong cách tiếp cận điều khiển cổng kết nối theo kiểu chủ/tớ như MGCP.

Mô hình giao thức MEGACO/H.248 có cấu trúc lệnh khá đơn giản, mềm dẻo trong thiết kế, cung cấp các ưu điểm nhằm giảm dung lượng mào đầu bản tin, giảm giá thành và độ phức tạp, có khả năng đáp ứng với rất nhiều ứng dụng khác nhau.

MEGACO có thể được sử dụng trên rất nhiều phương tiện mạng khác nhau, bao gồm cả mạng dựa trên IP và mạng dựa trên ATM.

Giao thức chuẩn MEGACO không bị ràng buộc với bất cứ cách tiếp cận điều khiển cuộc gọi mức ngang hàng cụ thể nào. Kiến trúc của MEGACO dựa trên lớp điều khiển cổng nối phương tiện MGC, lớp cổng nối phương tiện MG và chính giao thức MEGACO.

Lớp MGC bao gồm tất cả khả năng điều khiển cuộc gọi và các dịch vụ như chuyển tiếp cuộc gọi, truyền, hội nghị, giữ cuộc gọi. Lớp này cũng cung cấp bất cứ giao thức mức ngang hàng nào để tương tác với các MGC khác hoặc các thực thể ngang cấp. Lớp này quản lý tất cả các tương tác đặc tính và quản lý các tương tác với các thủ tục báo hiệu như báo hiệu số 7.

Lớp cổng nối phương tiện cung cấp các kết nối phương tiện từ mạng chuyển mạch gói tới mạng chuyển mạch gói, tương tác với những kết nối phương tiện này, điều khiển các thiết bị cổng kết nối. Lớp này được xem như là lớp ở mức cuộc gọi, đóng vai trò "tớ" trong kiến trúc chủ/tớ.

Do giao thức MEGACO hoàn toàn độc lập với các giao thức điều khiển cuộc gọi ở mức ngang hàng khác như SIP, H.323, các hệ thống khác nhau có thể được sử dụng ở mức điều khiển cuộc gọi với chi phí tối thiểu ở mức điều khiển cổng nối.

4.5.2. Mô tả giao thức MEGACO

MEGACO sử dụng mô hình tài nguyên đơn giản hiệu quả để mô tả các thực thể lô-gíc trong cổng nối phương tiện được thiết bị điều khiển cổng nối phương tiện. Khái niệm cơ bản dựa trên hai thuật ngữ Kết cuối (*Termination)* và Ngữ cảnh

(*Context*). Termination xác định các luồng hoặc các phương tiện tài nguyên, cung cấp các tín hiệu và phát đi các sự kiện, thuộc tính, thống kê. Tất cả các tín hiệu, sự kiện, thuộc tính, thống kê được định nghĩa trong các gói tin giao thức.

Cấu trúc lệnh trong MEGACO khá đơn giản nhưng hiệu quả và mềm dẻo. Trong MEGACO chỉ có 7 câu lệnh tất cả: *Add, Subtract, Modify, Move, Notify, AuditValue, ServiceChange*. Tất cả các câu lệnh trên đều hoạt động ở Termination theo một phương thức nhất quán.

Các câu lệnh giữa MGC và MG có thể dễ dàng nhóm với nhau thành một phiên truyền (*transaction*) sử dụng các qui tắc mềm dẻo, đơn giản. Do vậy dung lượng mào đầu bản tin có thể giảm một cách đáng kể.

Các câu lệnh sử dụng bộ mô tả (descriptor) trong cách tiếp cận của MEGACO để nhóm các thành phần dữ liệu liên quan. Điều này làm tăng tính mềm dẻo và khả năng phân tích dữ liệu. Chỉ có những bộ mô tả cần cho những hoạt động mới được gửi theo cùng các câu lệnh, do đó giảm dung lượng mào đầu.

Tất cả các tín hiệu và sự kiện được xem là xảy ra tại Termination và chúng cung cấp cơ chế để tương tác với các thực thể phía xa đại diện bởi Termination đó. Các tín hiệu và sự kiện được định nghĩa trong gói tin. Tín hiệu bao gồm âm phát đi, các thông báo và hiển thị số chủ gọi. Các sự kiện bao gồm tín hiệu nhấc hạ máy, số DTMF nhận được, xác nhận âm fax.

Thực thi giao thức MEGACO/H248 qua các điểm tham chiếu MSF

Trong cấu trúc hệ thống MSF (Multiservice Switching Forum), giao thức MEGACO/H.248 có thể hoạt động trên 3 điểm tham chiếu theo mô hình MSF khác nhau.

MEGACO/H.248 nhằm mục đích chuyển cổng kết nối H.323 vào trong MGC và MG. Do đó, để MSF sử dụng giao thức MEGACO/H.248, cần xác định sự chuyển đổi giữa MGC và MG và khối chức năng MSF khác. Sự chuyển đổi này cho phép xác định đúng vị trí giao thức MEGACO/H.248.

Chương V
QoS VÀ CÁC YẾU TỐ CHÍNH ẢNH HƯỞNG ĐẾN CHẤT LƯỢNG THOẠI IP

5.1. CHẤT LƯỢNG DỊCH VỤ IP

5.1.1. Thế nào là chất lượng dịch vụ

Chất lượng dịch vụ - QoS (*Quality of Service*) được xem như là lớp dịch vụ - CoS (*Class of Service*) và loại dịch vụ - ToS (*Type of Service*). Mục tiêu cơ bản của CoS và ToS là đạt được băng thông và thời gian trễ cần thiết cho một ứng dụng đặc biệt nào đó. CoS cho phép người quản trị có thể nhóm các luồng gói khác nhau, mỗi luồng có trễ và các nhu cầu băng thông riêng. Trường ToS trong tiêu đề của gói IP cho phép CoS hỗ trợ.

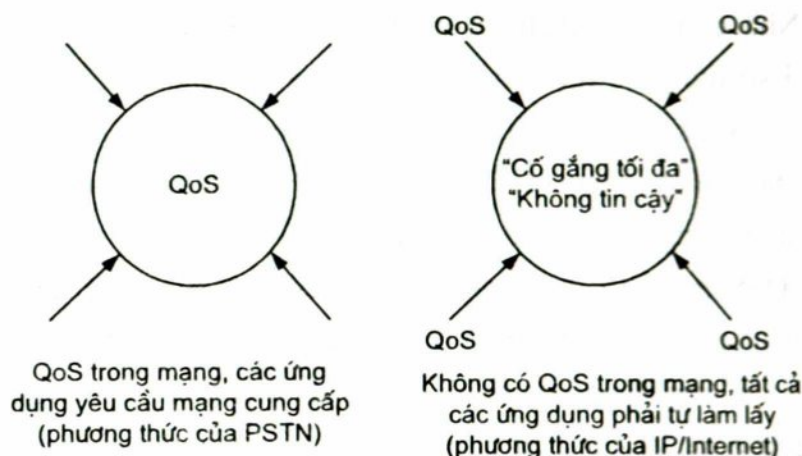
Nói một cách đơn giản nhất, chúng ta có thể coi chất lượng dịch vụ - QoS (*Quality of Service*) như những cách thức để có được những dịch vụ cung cấp dữ liệu ổn định và hiệu quả hay khả năng của mạng đảm bảo và duy trì các mức thực hiện nhất định cho mỗi ứng dụng theo các yêu cầu đã được chỉ rõ của mỗi người sử dụng.

QoS cho chúng ta biết khả năng của một phần tử mạng (có thể là một ứng dụng, là một trạm (host) hay một bộ định tuyến (router) để bảo đảm những yêu cầu nào đó về lưu lượng, về khả năng cung cấp dịch vụ…

Nhà cung cấp dịch vụ mạng đảm bảo QoS cung cấp cho người sử dụng và thực hiện các biện pháp để duy trì mức QoS khi điều kiện mạng bị thay đổi vì các nguyên nhân như: nghẽn, hỏng hóc thiết bị hay lỗi liên kết... QoS cần được cung cấp cho mỗi ứng dụng để người sử dụng có thể chạy ứng dụng đó và mức QoS mà ứng dụng đòi hỏi chỉ có thể được xác định bởi người sử dụng, bởi vì chỉ có người sử dụng mới có thể biết được chính xác ứng dụng của mình cần gì để hoạt động tốt. Người sử dụng tự động có thể biết được mạng cần cung cấp những gì cần thiết cho ứng dụng của mình, giống như các yêu cầu cho các ứng dụng phần mềm của máy tính cá nhân (ví dụ: cấu hình máy 128 MB RAM, 100 MB ổ cứng trống...).

Để cụ thể hơn, chúng ta hãy xem xét đến vấn đề QoS của một hệ thống mạng. Trong trường hợp này, QoS không liên quan đến một giải pháp tăng dải thông sẵn có của đường truyền trên mạng. Bởi vì thật phi lý khi đòi hỏi ở đường truyền một dải thông lớn hơn khả năng nó có thể cung cấp, mà QoS sẽ liên quan đến cách quản lý dải thông của đường truyền trước những yêu cầu của các ứng dụng cũng như liên quan đến cách thiết lập cấu hình mạng.

Về quan niệm thì QoS là một tài nguyên của mạng, hay đúng hơn là một nhóm tài nguyên mạng, ví dụ như băng thông và trễ. Quan niệm QoS là tài nguyên mạng như được chỉ ra ở hình 5.1. QoS ở "bên trong" mạng và mạng sẽ cung cấp QoS thích hợp cho các ứng dụng khi cần thiết.



Hình 5.1: Xây dựng mạng có QoS hay bổ sung QoS vào mạng

Các nhà cung cấp dịch vụ mạng đưa ra thông tin đặc tả về giá trị thực tế của các thông số QoS theo một trong hai cách sau. Với môi trường kênh ảo cố định (PVC- *Permanent Virtual Circuit*), các giá trị của các tham số QoS có thể chỉ đơn giản được ghi trên một mẩu giấy và trao cho đại diện của nhà cung cấp dịch vụ mạng. Fax, thư điện tử và thậm chí cả gọi điện để cung cấp cùng một thông tin về yêu cầu của khách hàng, miễn là khách hàng và nhà cung cấp dịch vụ thống nhất được với nhau về cách thức sử dụng để trao đổi thông tin về QoS. QoS có hiệu lực trên PVC khi PVC sẵn sàng. Với môi trường kênh ảo chuyển mạch (SVC: Switched Virtual Circuit), các giá trị của thông số QoS được gửi cho nhà cung cấp dịch vụ trong bản tin báo hiệu thiết lập cuộc gọi, nó là một phần của giao thức báo hiệu được sử dụng để cung cấp dịch vụ chuyển mạch trên mạng. Cả hai phương pháp đều có thể được sử dụng trong mọi cách kết hợp

trên mạng. Phương pháp PVC cho phép QoS được cung cấp rành rang hơn, trong khi phương pháp SVC đòi hỏi QoS trên một kết nối cho trước được thiết lập liên tục.

Nếu một mạng được tối ưu hoàn toàn cho một loại dịch vụ, thì người sử dụng ít phải xác định chi tiết các thông số QoS. Ví dụ, với mạng PSTN, được tối ưu cho thoại, không cần phải xác định băng thông hay trễ cần cho một cuộc gọi. Tất cả các cuộc gọi đều được đảm bảo QoS như đã được quy định trong các chuẩn liên quan cho điện thoại. Nếu mô-đem được sử dụng để truyền dữ liệu trên kết nối thoại thì các thông số được cung cấp bởi PSTN vẫn không thể bị vi phạm. PSTN thích hợp cho thoại do bản chất của nó, tuy nhiên nó lại không hoàn toàn phù hợp cho nhiều ứng dụng ngày nay, đặc biệt đối với khía cạnh băng thông. Tuy nhiên, cố định tất cả các thông số QoS trong mạng PSTN rất phù hợp cho thiết lập cuộc gọi và định tuyến khá đơn giản, hiệu quả và nhanh.

Thông thường, có nhiều mức QoS khác nhau cũng giống như là có nhiều ứng dụng vậy. Các ứng dụng lại thay đổi rất lớn thậm chí cả với những yêu cầu đơn giản về băng thông. Thoại số có thể yêu cầu bất kỳ tốc độ nào trong khoảng từ 8 đến 64 kbit/s. Video thì có thể chấp nhận được ở tốc độ 1,5 Mbit/s, nhưng sẽ tốt hơn nhiều nếu là 6 Mbit/s. Truy nhập Web và truyền tập tin sử dụng băng thông càng nhiều càng tốt trong phạm vi có thể, nhưng lại không cần liên tục... Tuy nhiên, băng thông trên PSTN và của mạng dữ liệu nhận được từ các đường thuê riêng dựa trên PSTN lại chỉ phục vụ giới hạn tại tốc độ 64 kbit/s. Một ứng dụng

cần 100 kbit/s thì hoặc là cố gắng hoạt động ở 64 kbit/s, hoặc là bỏ phí 28 kbit/s nếu sử dụng 128 kbit/s. Đây là mặt hạn chế của các mạng chuyển mạch kênh "tất cả băng thông tất cả thời gian".

Mạng chuyển mạch gói có thể chia băng thông ra thành nhiều phần rất thích hợp cho các ứng dụng dữ liệu bùng nổ, nhưng đó không phải là tất cả. Một mạng cần phải có khả năng cung cấp QoS yêu cầu cho mỗi ứng dụng, không cần biết băng thông cần thiết là cố định hay không. Khả năng về phía mạng cấp cho các ứng dụng các đảm bảo QoS, ví dụ như là đảm bảo băng thông, được xem là phân cấp QoS của mạng.

Phân cấp là một khía cạnh quan trọng của QoS. Phân cấp xác định các thông số QoS tốt đến mức nào mà người sử dụng có thể định rõ cho một ứng dụng cụ thể. Nếu mạng cung cấp QoS không đủ độ mịn, thì nó có thể giới hạn người sử dụng truy nhập vào mạng. Lấy một ví dụ đơn giản, xét một nhà cung cấp dịch vụ mạng thiết lập nhiều loại *lớp dịch vụ* cho các ứng dụng của người sử dụng. Có nhiều lúc lớp dịch vụ được dùng đồng nghĩa với QoS, nhưng ở đây sử dụng thuật ngữ lớp dịch vụ với nghĩa là một tập hợp của các giá trị thông số QoS có khả năng cung cấp cho một lớp, hay loại ứng dụng nào đó.

Nhà cung cấp dịch vụ có thể đưa ra một lớp dịch vụ thoại trên một mạng gói mà nó đảm bảo băng thông 64 kbit/s giữa các đầu cuối là trễ 100 ms với biến động trễ nhỏ hơn 10 ms. Nhưng nếu một ứng dụng thoại mới chỉ yêu cầu 8 kbit/s thôi thì sao? Hay thậm chí chỉ 4 kbit/s. Bởi vì người sử dụng được đảm bảo ở 64 kbit/s, nên lượng băng thông này nói chung phải được chia ra

từ toàn bộ băng thông trên mạng. Theo đó, người sử dụng trả cho phần băng thông không sử dụng và nhà cung cấp dịch vụ dự trữ băng thông để có thể dùng để cung cấp cho những người sử dụng khác.

Phân cấp QoS tốt hơn sẽ cho phép người sử dụng thoại, thậm chí trong cùng lớp dịch vụ, xác định băng thông họ yêu cầu chính xác hơn. Sự chính xác này muốn đạt được thì phải trả giá bằng độ phức tạp của mạng, đây là lý do chính cho việc giới hạn các giá trị thông số QoS và đặt ra các lớp dịch vụ trong giai đoạn đầu.

5.1.2. Tại sao chúng ta cần đến chất lượng dịch vụ cho mạng IP

Như chúng ta đã biết, giao thức cơ sở của Internet: giao thức IP cũng như cấu trúc của bản thân Internet đều dựa trên sự trao đổi các gói dữ liệu, các gói dữ liệu này, với phần địa chỉ nguồn và đích lưu trong tiêu đề (header), có thể tự lưu chuyển qua các bộ định tuyến liên kết giữa các mạng con mà không cần tới sự trợ giúp của phía gửi hay nhận chúng.

Tuy nhiên, việc đơn giản hóa hoạt động của mạng cũng có cái giá của nó. Bởi giao thức IP rất đơn giản, nên nó không có khả năng cung cấp nhiều dịch vụ. IP có thể cung cấp địa chỉ đích cho các gói tin dữ liệu cũng như có thể phân đoạn các gói tin để rồi ráp lại chúng sau đó, nhưng IP không thể cung cấp một cơ chế truyền dữ liệu đảm bảo tin cậy. Các bộ định tuyến có quyền hủy bỏ các gói tin IP mà không cần thông báo gì cho phía gửi cũng như phía nhận. Tất nhiên là IP sẽ dựa vào các giao thức truyền thông ở tầng trên nó, ví dụ TCP để có thể quản lý các gói tin. Tuy nhiên các giao thức này cũng chỉ bảo đảm rằng các gói

tin sẽ được gửi đi chứ không thể đảm bảo về tính thời gian thực của các gói tin đến hay về thông lượng dữ liệu. Như chúng ta đã biết, giao thức IP chỉ cung cấp được cái gọi là dịch vụ "best effort" (cố gắng tối đa), mà thực chất là dịch vụ truyền gói tin nhưng không quan tâm đến khi nào gói tin sẽ đến đích cũng như có bao nhiêu gói tin có thể đến đích.

Điều gì sẽ xảy ra nếu mạng không thành công trong việc đảm bảo và duy trì QoS chính xác cho một ứng dụng cho trước? Điều này tùy thuộc vào sự thỏa thuận giữa người sử dụng và nhà cung cấp dịch vụ trong trường hợp dịch vụ được quản lý bởi hợp đồng, hay là giữa nhà cung cấp dịch vụ và bộ phận điều chỉnh trong trường hợp dịch vụ bảng giá. Đảm bảo chất lượng cũng là phần quan trọng của hợp đồng cho các dịch vụ mạng giữa khách hàng (đại diện cho người sử dụng) và nhà cung cấp dịch vụ. Thông thường thì khách hàng phải trả tiền hàng tháng cho các dịch vụ. Đảm bảo chất lượng có thể thiết lập một hình thức bồi thường như giảm bớt giá tiền dịch vụ một tháng nếu nhà cung cấp dịch vụ không cung cấp đúng chất lượng yêu cầu trong tháng đó. Trong những trường hợp nghiêm trọng nếu vấn đề về mạng xảy ra trong toàn bộ tháng đó, khách hàng có thể nhận được dịch vụ miễn phí.

Đảm bảo chất lượng mạng trong một môi trường dịch vụ, hợp đồng thường được thể hiện theo hình thức thoả thuận mức dịch vụ (SLA: *Service Level Agreement*) được thiết lập giữa khách hàng và nhà cung cấp dịch vụ. SLA có thể là một phần của hợp đồng dịch vụ hay là một tiêu chuẩn độc lập hoàn toàn. SLA

đưa ra các yêu cầu của khách hàng và các hình thức bồi thường đối với nhà cung cấp trong trường hợp xảy ra sự cố. SLA cũng cung cấp một phương pháp thuận tiện cho khách hàng để so sánh các dịch vụ do các nhà cung cấp dịch vụ khác nhau đưa ra.

Vậy trong tất cả những điều đã nêu về phân cấp QoS, đảm bảo chất lượng và SLA, điều nào phải được thực hiện với VoIP và thoại Internet. Vấn đề ở đây là trong khi đảm bảo và điều chỉnh QoS trở thành một vấn đề được xem xét tích cực giữa các nhà cung cấp dịch vụ mạng công cộng, Internet nhìn chung vẫn tương đối không bị ảnh hưởng trong vấn đề QoS. Một trong những lý do mà bản chất Internet định hướng IP là một mạng "cố gắng tối đa", do đó "không tin cậy" khi nó tiến tới đảm bảo về QoS. Tuy nhiên, nếu tất cả các nhà cung cấp dịch vụ Internet (Internet Service Provider - ISP) muốn quan tâm tới QoS, thì vẫn không dễ gì thêm QoS vào một mạng IP tại lớp IP.

Phương pháp tốt nhất để các nhà cung cấp dịch vụ Internet có thể đạt tới đảm bảo QoS hay thỏa thuận kết nối mức dịch vụ (SLA) giữa khách hàng và ISP là với dịch vụ mạng IP *được quản lý*. Thuật ngữ *được quản lý* ở đây là bất cứ cái gì mà nhà cung cấp dịch vụ quản lý thay mặt cho khách hàng. Vậy thì cái gì đang thực sự được quản lý trên mạng IP? Đó là QoS mà mạng IP cung cấp. Điều này được thực hiện bằng cách cách ly các bộ định tuyến, các liên kết.... sử dụng để cung cấp dịch vụ IP cho một khách hàng cụ thể và sử dụng các tài nguyên này trên một nền tảng dành riêng một phần phục vụ cho mình khách hàng đó. Trong vài trường hợp, các bộ định tuyến và các liên kết vẫn

được chia sẻ, nhưng chỉ giữa những khách hàng chung vốn có hợp đồng cho quản lý các dịch vụ IP.

Hầu hết các nhà cung cấp dịch vụ Internet (ISP) lớn đưa ra cả kết nối Internet công cộng dùng chung và dịch vụ IP được quản lý. Phần IP được quản lý của các ISP thường sử dụng để kết nối chỉ với các vị trí được điều khiển bởi khách hàng. Không ai có thể dễ dàng đảm bảo băng thông hay bất kỳ một thông số QoS nào khác trên mạng Internet công cộng, nó cơ bản bao gồm các "đám mây" của các ISP liên kết với băng thông và tài nguyên biến đổi trong phạm vi lớn. Chỉ có thể bằng cách giới hạn các kết nối đến một ISP thì ISP mới có thể đưa ra thực tế một dịch vụ IP được quản lý. Các liên kết đến một mạng Internet công cộng toàn cầu có thể được cung cấp như một phần của dịch vụ IP được quản lý, nhưng tất nhiên là tất cả các đảm bảo QoS không xuất hiện trên phần này của mạng. Tuy nhiên, liên kết giới hạn được đặc trưng bởi các dịch vụ mạng IP được quản lý này vẫn có thể được sử dụng để đem lại lợi ích cho khách hàng. Ví dụ, mạng riêng ảo (VPN: *Virtual Private Network*) thực sự được lợi từ các đảm bảo QoS và các kết nối hạn chế vị trí – và đây chính là dấu xác nhận chất lượng của các dịch vụ mạng IP được quản lý.

Vấn đề ở đây là ngày càng có nhiều ứng dụng như thoại và video hoạt động trên Internet và Web toàn cầu, do đó các đảm bảo QoS không chỉ đơn thuần là xa xỉ mà thực sự cần thiết. Mặc dù vậy, Internet ngày nay tuy cung cấp liên kết toàn cầu nhưng lại có rất ít các đảm bảo QoS, có chăng chỉ là các ngoại lệ của các dịch vụ mạng IP được quản lý.

Hạn chế này chưa phải là một vấn đề khó khăn khi chúng ta xét đến những dịch vụ Internet truyền thống như Web, email, FTP... Nhưng với những dịch vụ mới đòi hỏi thông lượng cao và độ trễ thấp, ví dụ như các dịch vụ truyền âm thanh và hình ảnh, thì đây quả là một trở ngại lớn.

Không giống như những công nghệ truyền thông sử dụng các kênh ảo (virtual circuit) như ATM và chuyển tiếp khung (Frame-Relay), IP không quan tâm nhiều đến việc phân phối tài nguyên mạng. Một mặt, điều này cho phép tận dụng dải thông hiệu quả nhưng mặt khác, nó sẽ khiến cho vấn đề chia sẻ đường truyền trở nên phức tạp hơn nhiều. Điều này có thể gây ra những lỗi không thể đoán trước được. Để có thể khắc phục được những lỗi này, chúng ta cần áp dụng những mức độ bảo đảm riêng cho chất lượng mạng. Và như đã trình bày cần cung cấp thêm QoS vào trong mạng IP để nó có thể đáp ứng tốt hơn cho nhiều ứng dụng, nhất là các ứng dụng đa phương tiện.

5.2. CÁC THÔNG SỐ QoS

Trước đây, lý do chính tại sao thoại IP chưa được sử dụng là do giao thức TCP/IP không hỗ trợ tốt cho các ứng dụng thời gian thực, giao thức UDP/IP đảm bảo được độ trễ thấp nhưng độ tin cậy không cao. Băng thông và trễ trong mạng IP quá lớn cũng là vấn đề không phù hợp với các ứng dụng thời gian thực như thoại. Các giao thức và kỹ thuật mã hóa được sử dụng sau này trong mạng IP đã giải quyết được phần nào các khó khăn trên, nhưng để chất lượng thoại IP đảm bảo, một số vấn đề về QoS cần phải giải quyết tốt hơn.

Tính đến thời điểm này các khái niệm của các thông số QoS mới chỉ được thảo luận bằng những thuật ngữ chung nhất mà vẫn chưa có định nghĩa rõ ràng các thông số QoS. Tuy nhiên, hãy suy nghĩ đến tầm quan trọng của QoS trong VoIP và điện thoại Internet.

Phần này định nghĩa QoS như là đặc tả kỹ thuật của các giá trị trên nền tảng ứng dụng với ứng dụng đối với mỗi một trong sáu thông số. Sáu thông số, cùng với các giá trị ví dụ, được liệt kê trong bảng 5.1.

Bảng 5.1: Sáu thông số của QoS

Thông số QoS	**Các giá trị ví dụ**
Băng thông (nhỏ nhất)	64 kbit/s, 1,5 Mbit/s, 45 Mbit/s
Trễ (lớn nhất)	50 ms trễ vòng, 150 ms trễ vòng
Biến động trễ (jitter)	10% của trễ lớn nhất, 5 ms biến động
Mất gói (các ảnh hưởng của lỗi)	1 trong 1000 gói chưa chuyển giao
Độ khả dụng (độ tin cậy)	99,99%
Bảo mật	Mã hóa và nhận thực trên tất cả các luồng lưu lượng

5.2.1 Băng thông (bandwidth)

Khá nhiều vấn đề về QoS đều bắt đầu và kết thúc với băng thông. Băng thông rõ ràng là thông số quan trọng nhất của QoS, nhưng nó cũng không thể là thông số duy nhất để xác định QoS. Trong nhiều phần của mạng, băng thông vẫn là "mặt hàng xa xỉ", mặc dù có các hứa hẹn "băng thông gần như không giới hạn" trên sợi quang. Không may rằng, những thất bại với các cuộc thử nghiệm hầu như không có sự tham gia của điện và

"phần cứng gần như tự cấu hình" đã tạo ra các hoài nghi về khả năng sớm đạt được băng thông như yêu cầu.

Băng thông chỉ đơn giản là thước đo số lượng bit trên giây mà mạng sẵn sàng cung cấp cho các ứng dụng. Các ứng dụng bùng nổ (bursty) trên mạng chuyển mạch gói có thể chiếm tất cả băng thông của mạng nếu không có ứng dụng nào khác cùng bùng nổ với nó. Khi điều này xảy ra, các bùng nổ phải được đệm lại và xếp hàng chờ truyền đi, do đó tạo ra trễ trên mạng.

Khi được sử dụng như là một thông số QoS, băng thông là yếu tố tối thiểu mà một ứng dụng cần để hoạt động. Ví dụ, thoại PCM 64 kbit/s cần băng thông là 64 kbit/s. Điều này không tạo ra khác biệt khi mạng đường trục có kết nối 45 Mbit/s giữa các nút mạng lớn. Băng thông cần thiết được xác định bởi băng thông nhỏ nhất sẵn có trên mạng. Nếu truy nhập mạng thông qua một mô-đem V.34 hỗ trợ chỉ 33,6 kbit/s, thì mạng đường trục dù tốc độ là 45 Mbit/s thì ứng dụng thoại 64 kbit/s vẫn không hoạt động được. Băng thông QoS nhỏ nhất phải sẵn sàng tại tất cả các điểm giữa các người sử dụng.

Các ứng dụng dữ liệu được lợi nhất từ việc đạt được băng thông cao hơn. Điều này được gọi là các ứng dụng giới hạn băng thông, bởi vì hiệu quả của ứng dụng dữ liệu trực tiếp liên quan tới lượng nhỏ nhất của băng thông sẵn sàng trên mạng. Mặt khác, các ứng dụng thoại như thoại PCM 64 kbit/s được gọi là các ứng dụng giới hạn trễ. Thoại 64 kbit/s sẽ không hoạt động tốt hơn là mấy nếu có băng thông 128 kbit/s. Loại thoại này phụ thuộc hoàn toàn vào thông số QoS trễ của mạng để có thể hoạt động đúng đắn.

Gói trong mạng chuyển mạch gói yêu cầu tiêu đề gói, chiều dài gói và trường tổng kiểm tra (checksum). Sử dụng RTP/UDP/IPv4 tiêu đề gói là 40 byte, nếu sử dụng RTP/UDP/IPv6 thì tiêu đề là 60 byte. Số lượng dữ liệu thoại chứa trong một gói phụ thuộc vào bộ mã hóa thoại được sử dụng và số của khung chứa trong mỗi gói. Ví dụ 20 ms của thoại được mã hóa trong mỗi gói IPv4. Nếu sử dụng bộ mã hóa CS-CELP khung có chiều dài 10 ms sử dụng 10 byte, hai khung tương ứng là 20 byte và như vậy tỷ lệ tiêu đề gói là 67%.

Bình thường, một kênh thoại tương tự được biến đổi thành một kênh PCM cơ sở có tốc độ 64 kbit/s. Kỹ thuật PCM (chuẩn G.711) sử dụng trong mạng thoại truyền thống đảm bảo chất lượng âm khá trung thực song lượng dải thông sử dụng còn lớn. Điều này không phù hợp khi sử dụng trong mạng IP. Vấn đề đặt ra là giảm băng thông hơn nữa mà vẫn đảm bảo chất lượng dịch vụ để phù hợp với mạng VoIP, từ đó xuất hiện một số kỹ thuật mã hóa và nén tín hiệu thoại tốc độ thấp cụ thể như GSM G.723.1, G.729 và G.729A. PCM và ADPCM đều là kỹ thuật mã hóa theo dạng sóng lợi dụng các đặc tính của bản thân dạng sóng. Kỹ thuật nén mới (mã hóa nguồn) được phát triển cách đây 10-15 năm dựa vào sự nhận biết các đặc tính nguồn của tín hiệu thoại phát ra. Kỹ thuật này sử dụng thủ tục xử lý tín hiệu và nén thoại bằng việc chỉ gửi đi thông tin ở dạng các tham số đã được đơn giản hóa về việc kích tín hiệu nguồn dạng của giọng nói do vậy đòi hỏi ít băng thông hơn.

5.2.2 Trễ (delay)

Băng thông và trễ của mạng, có quan hệ với nhau và có thể tính toán tại nhiều vị trí trong mạng, thậm chí từ đầu cuối tới đầu

cuối. Thông tin truyền đi dưới dạng một chuỗi các khung truyền (gói IP cũng có thể được sử dụng cho mục đích này). Khoảng thời gian trôi qua từ khi bit đầu tiên của một khung đi vào mạng cho đến khi bit đầu tiên rời khỏi mạng gọi là trễ. Vì con đường của khung qua cả bộ chuyển mạch và bộ định tuyến, nên trễ có thể biến đổi, có các giá trị lớn nhất, nhỏ nhất, trung bình, độ lệch chuẩn...

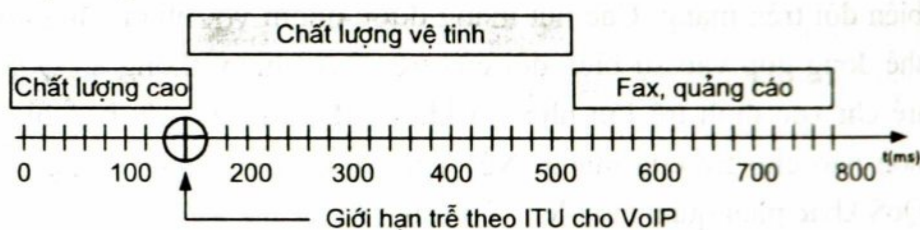
Băng thông được định nghĩa là số bit của một khung chia cho thời gian kể từ khi bit đầu tiên rời khỏi mạng cho đến khi bit cuối cùng rời mạng. Trên thực tế, đây chỉ là một trong nhiều cách đo có thể. Vì các khung có đường đi từ liên kết truy nhập tới mạng đường trục, nên băng thông mà khung được truyền đi có thể biến đổi đáng kể.

Các mạng chuyển mạch gói cung cấp cho các ứng dụng yêu cầu băng thông biến đổi phụ thuộc vào hoạt động và bùng nổ của ứng dụng. Băng thông biến đổi này có nghĩa là trễ cũng có thể biến đổi trên mạng. Các nút mạng được nhóm với nhau cũng có thể đóng góp vào sự biến đổi của trễ. Tuy nhiên, thông số QoS trễ chỉ xác định trễ lớn nhất và không đặt bất kỳ giới hạn nhỏ hơn nào cho trễ của mạng. Nếu cần trễ ổn định, một thông số QoS khác phải quan tâm đến yêu cầu này.

Trong trễ tổng đầu cuối - đầu cuối của gói bao gồm trễ xử lý, trễ mạng và trễ bộ đệm. Trễ xử lý bao gồm trễ mã hóa và trễ đóng gói (nó tuỳ thuộc vào bộ mã hóa). Trễ mã hóa là thời gian xử lý cần thiết để mã hóa và giải mã tín hiệu. Trễ đóng gói bao gồm khoảng thời gian của đoạn tín hiệu thoại đóng gói, thông thường từ 10 đến 40 ms. Khi thoại được đóng gói sẽ sinh ra trễ

thuật toán do nhất thiết phải lấy mẫu tương ứng với một khung trước khi đưa vào xử lý. Trễ thuật toán liên quan đến cỡ khung và thay đổi từ vài phần ms đến vài chục ms. Trễ xử lý phụ thuộc vào thuật toán nén và tốc độ bộ xử lý.

Thực tế bộ mã hóa và giải mã là các thiết bị vật lý độc lập, do đó trễ mã hóa và trễ giải mã được xem xét riêng. Ngoài ra, còn có trễ khi gói đi từ thiết bị phát đến thiết bị thu, trễ này gọi là trễ truyền dẫn (mạng), nó phụ thuộc vào khoảng cách, dung lượng và trạng thái của liên kết trong mạng. Trễ mạng thường thay đổi và khó biết trước, phụ thuộc nhiều vào tải trên mạng, hiện tượng trễ mạng thay đổi được gọi là biến động trễ (jitter) và thường giải quyết bằng bộ đệm biến động trễ ở đầu thu. Trễ đầu cuối - đầu cuối là một tham số quan trọng khi xem xét các ứng dụng thông tin yêu cầu thời gian thực như thoại. Nếu trễ này lớn quá thì quá trình đàm thoại sẽ không thực hiện được. Theo chuẩn ITU G.114 thì trễ đầu cuối - đầu cuối nhỏ hơn 150 ms là chấp nhận được.



Hình 5.2: Trễ đầu cuối - đầu cuối

5.2.3 Biến động trễ (Jitter)

Thông số QoS biến động trễ thiết lập giới hạn lên lượng biến đổi của trễ mà một ứng dụng có thể gặp trên mạng. Một cách đúng đắn hơn thì "jitter" được xem là biến động trễ, bởi vì thuật

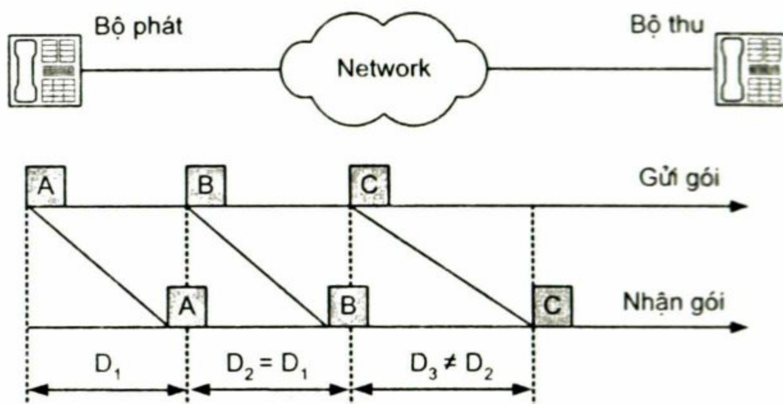
ngữ "jitter" cũng được sử dụng trong mạng với nghĩa sự khác biệt thời gian mức thấp trong kỹ thuật mã đường dây. Tuy nhiên, sử dụng thuật ngữ "jitter" đồng nghĩa với biến động trễ cũng là phổ biến và ngữ cảnh sẽ luôn phân biệt nghĩa nào đang được đề cập. Biến động trễ không đặt một giới hạn nào cho giá trị tuyệt đối của trễ, nó có thể tương đối thấp hoặc cao phụ thuộc vào giá trị của thông số trễ.

Biến động trễ theo lý thuyết có thể là một giá trị thông số QoS mạng tương đối hay tuyệt đối. Ví dụ, nếu trễ mạng cho một ứng dụng được thiết lập là 100 ms, biến động trễ có thể đặt là cộng hay trừ 10 phần trăm của giá trị này. Theo đó, nếu mạng có trễ trong khoảng 90 đến 110 ms thì vẫn đạt được yêu cầu về biến động trễ (trong trường hợp này, rõ ràng là trễ không phải là lớn nhất). Nếu trễ là 200 ms, thì 10 phần trăm giá trị biến động trễ sẽ cho phép bất kỳ trễ nào trong khoảng 180 đến 220 ms. Mặt khác, biến động trễ tuyệt đối giới hạn cộng trừ 5 ms sẽ giới hạn biến động trễ trong các ví dụ trên trong khoảng từ 95 tới 105 ms và từ 195 tới 205 ms.

Các ứng dụng nhạy cảm nhất đối với giới hạn của biến động trễ là các ứng dụng thời gian thực như thoại hay video. Nhưng đối với các trang Web hay với truyền tập tin qua mạng thì lại ít quan tâm hơn đến biến động trễ. Internet, là gốc của mạng dữ liệu, có ít khuyến nghị về biến động trễ. Các biến đổi của trễ tiếp tục là vấn đề gây khó khăn nhất gặp phải đối với các ứng dụng video và thoại dựa trên Internet.

Trên hình 5.3 cho thấy khoảng thời gian từ khi gửi gói A và B đến khi nhận được chúng là bằng nhau ($D_1 = D_2$), nhưng

khoảng thời gian từ khi gửi gói C đến khi nhận được lại khác ($D_3 \neq D_2$). Đó chính là biến động trễ.



Hình 5.3: Sự biến động trong thời gian đến của gói

Trong trễ tổng đầu cuối - đầu cuối của gói bao gồm trễ xử lý, trễ mạng và trễ bộ đệm thì có tham số rất biến động đó là trễ mạng. Nó là thời gian để lưu chuyển gói thoại từ đầu phát đến đầu thu và nó phụ thuộc rất nhiều vào điều kiện mạng. Khoảng thời gian này với mỗi gói là khác nhau được gọi là biến động trễ. Đó là do cơ chế của mạng, hai gói được gửi từ cùng một đầu cuối đến cùng một đích nhưng lại đi qua các tuyến khác nhau dẫn tới trễ khác nhau thậm chí còn không đến theo thứ tự như khi phát nữa. Để giải quyết vấn đề này, bộ thu sử dụng bộ đệm để hấp thu biến động trễ, tức là các gói đến sẽ không được đọc ngay mà lưu trong bộ đệm và được đọc ra theo theo khoảng thời gian đều hoặc theo chương trình đọc ra xác định.

5.2.4 Mất gói (packet loss)

Mất gói gây ra mất thông tin, mất thông tin là một thông số QoS không được đề cập thường xuyên như là băng thông và trễ,

đặc biệt đối với mạng Internet. Đó là vì bản chất tự nhiên được thừa nhận của mạng Internet là "cố gắng tối đa". Nếu các gói IP không đến được đích thì Internet không hề bị đổ lỗi vì đã làm mất chúng. Điều này không có nghĩa là ứng dụng sẽ tất yếu bị lỗi, bởi vì nếu các thông tin bị mất vẫn cần thiết đối với ứng dụng thì nó sẽ phải tự yêu cầu bên gửi gửi lại bản sao của thông tin bị mất. Bản thân mạng không hỗ trợ vấn đề này, bởi vì bản sao của thông tin bị mất không được lưu lại tại bất cứ nút nào của mạng.

Tại sao các mạng, không chỉ Internet, lại mất thông tin? Thực sự là có nhiều lý do, nhưng hầu hết là do các ảnh hưởng của lỗi trên mạng. Ví dụ, nếu một kết nối bị hỏng, thì tất cả các bit đang truyền trên liên kết này sẽ không và không thể tới được đích. Nếu một nút mạng, ví dụ như bộ định tuyến hỏng, thì tất cả các bit hiện đang ở trong bộ đệm và đang được xử lý bởi nút đó sẽ biến mất không để lại dấu vết. Do những loại hư hỏng này trên mạng có thể xảy ra bất cứ lúc nào, nên việc một số gói tin bị mất do lỗi trên mạng là không thể tránh khỏi.

Tác động của mất thông tin là tùy thuộc vào ứng dụng. Điều khiển lỗi trên mạng là một quá trình gồm hai bước, mà bước đầu tiên là xác định lỗi. Bước thứ hai là khắc phục lỗi, nó có thể đơn giản là bên gửi truyền lại đơn vị bị mất thông tin. Một vài ứng dụng, đặc biệt là các ứng dụng thời gian thực, không thể đạt hiệu quả khắc phục lỗi bằng cách gửi lại đơn vị tin bị lỗi. Các ứng dụng không phải thời gian thực thì thích hợp hơn đối với cách truyền lại thông tin bị lỗi, tuy nhiên cũng có một số ngoại lệ (ví dụ như các hệ thống quân sự tấn công mục tiêu trên không không thể sử dụng hiệu quả với cách khắc phục lỗi bằng truyền lại).

Vì những lý do này, thông số QoS mất thông tin không những nên định rõ một giới hạn trên đối với ảnh hưởng của lỗi mà còn nên cho phép người sử dụng xác định xem có lựa chọn cách sửa lỗi bằng truyền lại hay không. Tuy nhiên, hầu hết các mạng (đặc biệt là mạng IP) chỉ cung cấp phương tiện vận chuyển thụ động, còn xác định lỗi, khắc phục lỗi thường được để lại cho ứng dụng của người sử dụng.

Mạng IP hoạt động trên theo cơ chế "cố gắng tối đa" nên độ tin cậy không cao, mất gói thường xảy ra, đôi khi xảy ra nghẽn mạng và các gói phải nằm lâu trong hàng đợi của các nút mạng, xấu hơn là có thể rớt do tràn bộ đệm hoặc do độ ưu tiên của gói thấp bị bộ định tuyến loại bỏ. Để truyền tin cậy dữ liệu trong mạng IP người ta phải sử dụng cơ chế phát lại để truyền lại các gói bị mất, nó được thực hiện ở lớp giao vận. Cơ chế phát lại được kích hoạt khi đầu cuối nhận gửi bản tin thông báo mất gói. Tuy nhiên, việc phát lại có nhược điểm lớn là trễ. Đối với các ứng dụng thời gian thực như thoại thì cơ chế phát lại không thể áp dụng được.

Gói thực chất được xác định là mất khi trễ đầu cuối - đầu cuối của gói vượt quá quy định. Như vậy, gói mất là gói không bao giờ đến đích (thời gian trễ bằng vô cùng) nhưng cũng có thể là gói mặc dù đến đích nhưng sau thời điểm quy định đọc ra của chúng (trường hợp này gọi là mất do trễ). Số của gói mất sẽ giảm khi tăng trễ bộ đệm nghĩa là tăng thời gian trễ cho phép của gói. Tuy nhiên, tăng trễ bộ đệm đồng nghĩa với việc tăng trễ đầu cuối - đầu cuối (mà điều này lại ảnh hưởng đến chất lượng, đến độ cảm nhận của người nghe). Do đó, cần phải có sự cân nhắc thỏa đáng giữa hai tham số này.

Mất gói làm gián đoạn các đoạn thoại, gây nên các âm như click, mất tiếng (trong khoảng rất ngắn) hoặc mất âm làm cho đoạn thoại không hiểu được. Vì mất gói là điều không thể tránh khỏi, lưu lượng thoại cho phép có thể mất một số gói, nhưng nếu tỷ lệ mất gói này vượt quá 5% thì nó sẽ ảnh hưởng đáng kể tới chất lượng thoại. Do vậy cần phải có các biện pháp để giảm ảnh hưởng của mất gói đến chất lượng thoại. Đó là kỹ thuật che dấu mất gói hay rộng hơn là khôi phục mất gói, nghĩa là làm cho người nghe không cảm nhận thấy khó chịu do các tác động của mất gói gây nên như trình bày ở trên.

5.2.5 Độ tin cậy (Reliability)

Các mạng tồn tại để phục vụ người sử dụng. Tuy nhiên, mạng cần có biện pháp bảo dưỡng phòng ngừa nếu các tình huống hỏng hóc tiềm tàng được phát hiện và dự đoán trước. Một chiến lược đúng đắn bằng cách định kỳ tạm thời tách các thiết bị ra khỏi mạng để thực hiện các công việc bảo dưỡng và chuẩn đoán trong một khoảng thời gian ngắn có thể giảm thời gian ngừng hoạt động do hỏng hóc. Tất nhiên, thậm chí với một biện pháp bảo dưỡng hoàn hảo nhất cũng không thể tránh được các lỗi không tiên đoán trước và các lỗi nghiêm trọng của kết nối và thiết bị theo thời gian.

Thường thì mạng PSTN có quy trình thời gian và bảo dưỡng nghiêm khắc hơn nhiều mạng dữ liệu. PSTN phải có khả năng truyền các cuộc gọi cả ngày lẫn đêm và tất cả các ngày trong năm. Có những khoảng thời gian chỉ có rất ít cuộc gọi, như khoảng từ 3 đến 4 giờ sáng, nhưng lại có cuộc gọi hầu như tất cả

các khoảng thời gian. Đương nhiên, phải có những nguyên tắc để bảo dưỡng phòng ngừa với mạng PSTN. Một số hoạt động có thể thực hiện lúc lưu lượng được biết trước là tạm lắng, và một số hoạt động có thể không bao giờ được thực hiện trong các giờ hoặc ngày bận...

Mạng dữ liệu thực hiện công việc đó dễ hơn. Hầu hết mạng dữ liệu dành cho kinh doanh, và do đó hoạt động trong những giờ kinh doanh, thường là từ 8 giờ sáng đến 5 giờ chiều, từ thứ Hai đến thứ Sáu. Hoạt động bổ trợ có thể thực hiện "ngoài giờ", và một tập kiểm tra đầy đủ với mục đích phát hiện ra các vấn đề có thể chạy trong ngày nghỉ.

Internet và Web đã thay đổi tất cả. Mọi mạng toàn cầu phải giải quyết vấn đề rằng thực sự có một số người luôn cố gắng truy nhập vào mạng tại một số địa điểm. Và thậm chí Internet có thể có ích ở nhà vào 10 giờ tối hơn là ở cơ quan vào 2 giờ chiều.

Tuy nhiên, nếu người sử dụng nhận thức rõ rằng họ không thể có mạng như mong muốn trong tất cả thời gian, bao nhiêu thì đủ? Và khi nào hỏng hóc xảy ra, dịch vụ phải được khôi phục nhanh đến mức nào? Cả hai là khía cạnh chủ yếu của thông số QoS *độ khả dụng*, hay *độ tin cậy của mạng*.

Một năm có 60 x 60 x 24 x 365, hay 31.536.000 giây. Giả thiết một mạng khả dụng 99 phần trăm thời gian. Điều này cho phép một nhà cung cấp dịch vụ có 315.360 giây, hay 87,6 giờ mạng không hoạt động trong một năm. Khoảng thời gian này là tương đối lớn. Giá trị 99,99% sẽ tốt hơn nhiều, và thời gian

ngừng hoạt động của mạng giảm xuống chỉ còn khoảng 50 phút trong một năm. Tất nhiên nhà cung cấp dịch vụ cần nhiều cơ chế dự phòng và khắc phục lỗi hơn để đạt được điều này. Bảng cho thấy phần trăm sẵn sàng được biểu diễn dưới dạng thời gian ngừng hoạt động hàng năm.

Bảng 5.2: Tính sẵn sàng của mạng và thời gian ngừng hoạt động

Tính sẵn sàng của mạng	Tổng thời gian ngừng hoạt động trong một năm
99%	3,65 ngày
99,5%	1,825 ngày
99,9%	8,76 giờ
99,95%	4,38 giờ
99,99%	52,56 phút
99,995%	26,28 phút
99,999%	5,25 phút

Ngày nay, thông số QoS khả dụng của mạng thường vào khoảng 99,995%, hay khoảng 26 phút ngừng hoạt động trong một năm, kết nối khôi phục nhỏ hơn 4 giờ. Cũng có sự khác nhau giữa độ khả dụng và độ tin của mạng từ góc nhìn của từng người sử dụng và từ góc nhìn mạng tổng thể. Thông số QoS khả dụng thường được quy cho mỗi vị trí hoặc liên kết riêng lẻ. Một người sử dụng khó tính có thể than phiền rằng một liên kết chỉ sẵn sàng 99,7% trong tháng sẽ được nhắc nhở rằng 99,99% sẵn sàng như được quảng cáo và hứa hẹn là áp dụng cho toàn bộ mạng.

5.2.6 Bảo mật (Security)

Bảo mật là một thông số mới trong danh sách QoS, nhưng lại là một thông số quan trọng. Thực tế, trong một số trường hợp độ bảo mật có thể được xét ngay sau băng thông. Gần đây, sự do đe dọa mạnh mẽ của các tin tặc và sự lan tràn của vi-rút trên mạng Internet toàn cầu đã làm cho bảo mật trở thành vấn đề hàng đầu.

Hầu hết bảo mật liên quan tới các vấn đề như tính riêng tư, sự tin cẩn, xác nhận khách và chủ. Chương này sẽ không đi sâu vào chi tiết về vấn đề mạng và bảo mật, mà chỉ nhắc nhở rằng các vấn đề liên quan đến bảo mật thường được gắn với một vài hình thức của phương pháp mật mã, như mã hóa và giải mã. Các phương pháp mật mã cũng được sử dụng trên mạng cho việc nhận thực (authentication), nhưng những phương pháp này thường không liên quan chút nào đến vấn đề giải mã.

Một cách ngắn ngọn, riêng tư và bí mật có liên quan tới các kỹ thuật mã hóa riêng tư hay công khai. Việc xác nhận tính hợp lệ của khách hàng ("bạn đúng là người như bạn đã nói") thường được quy định bởi một mật khẩu đơn giản, nếu sử dụng chữ ký số thì phức tạp hơn, và thậm chí còn phức tạp hơn nữa nếu sử dụng các hệ thống sinh trắc học như kiểm tra võng mạc. Việc xác nhận tính hợp lệ của máy phục vụ thường được quy định bởi một chứng chỉ số được đưa ra bởi nhà cấp chứng chỉ và được quản lý bởi một nhà quản lý đăng ký.

Toàn bộ kiến trúc đều xuất phát từ việc bổ sung thêm tính riêng tư hoặc bí mật và sự xác nhận hoặc nhận thực cho mạng Internet. Giao thức bảo mật chính thức cho IP, gọi là IPSec, đang trở thành một kiến trúc cơ bản để cung cấp thương mại

điện tử trên Internet và ngăn ngừa gian lận trong môi trường VoIP. Mạng Internet công cộng toàn cầu, thường xuyên bị coi là thiếu bảo mật nhất, đã đưa vấn đề về bảo mật trở thành một phần của IP ngay từ khi bắt đầu. Một bit trong trường *loại dịch vụ* (ToS: Type of Service) trong phần tiêu đề gói IP được đặt riêng cho ứng dụng để có thể bắt buộc bảo mật khi chuyển mạch gói. Tuy nhiên lại nảy sinh một vấn đề là không có sự thống nhất giữa các nhà sản xuất bộ định tuyến khi sử dụng trường ToS.

Người sử dụng và ứng dụng có thể thêm phần bảo mật của riêng mình vào mạng và trong thực tế, cách này đã được thực hiện trong nhiều năm. Nếu có chút nào bảo mật mạng, thì nó thường dưới dạng một mật khẩu truy nhập vào mạng. Các mạng ngày nay cần một cơ chế bảo mật gắn liền với nó, chứ không phải thêm vào một cách bừa bãi bởi các ứng dụng. Nếu không khả năng kết hợp của tương tác khách-chủ gồm cả bảo mật sẽ trở thành một cơn ác mộng.

Một thông số QoS bảo mật điển hình có thể là "mã hóa và nhận thực đòi hỏi trên tất cả các luồng lưu lượng". Nếu có lựa chọn, thì truyền dữ liệu có thể chỉ cần mã hóa, và kết nối điện thoại Internet có thể chỉ cần nhận thực để ngăn gian lận. Ngày nay, tầm quan trọng của bảo mật như một thông số QoS là rất lớn, không thể đánh giá hết được.

5.3. CÁC YẾU TỐ CHÍNH ẢNH HƯỞNG ĐẾN CHẤT LƯỢNG THOẠI VÀ CÁC GIẢI PHÁP

5.3.1 Vai trò quan trọng của băng thông

Một số danh sách các thông số QoS của mạng có thể rất ngắn. Một số chỉ gồm băng thông, trễ và biến động trễ là các

biến hợp pháp mà người sử dụng và ứng dụng có thể định rõ trên mạng. Phần còn lại chỉ là những gì mà mạng cấp cho người sử dụng, do đó người sử dụng không cần thiết phải xác định các giá trị cho bảo mật hay bất kỳ một thông số nào khác.

Một số danh sách chỉ có một mục - băng thông. Bây giờ thậm chí trễ và biến động trễ chỉ là những gì mà mạng có thể đưa ra, không phải những gì mà người sử dụng và ứng dụng cần hay mong muốn. Cho trước băng thông đủ dùng thì tất cả các thông số QoS liên quan khác sẽ trở nên mờ nhạt. Đây là một lý lẽ có đầy sức thuyết phục. Ảnh hưởng của việc tăng băng thông đến giảm trễ đã được đề cập ở phần trước. Người sử dụng có thể chỉ cần chọn trễ, biến động trễ... nào mà họ cần và một băng thông lớn sẽ giải quyết tất cả. Không cần phải định rõ bất kì một thông số QoS nào khác trừ băng thông.

Ví dụ, băng thông lớn cho phép các gói tin đến đích hầu như lập tức, do đó tối thiểu trễ. Không bao giờ có hàng đợi nên biến động trễ cũng biến mất. Lỗi không còn là vấn đề bởi vì người sử dụng có thể gửi mọi thứ nhiều lần trong cùng một thời điểm và việc sử dụng nhiều bộ định tuyến sẽ giảm thiểu lỗi. Bảo mật không là vấn đề gì vì người sử dụng có thể thay đổi định tuyến các bit nếu họ thích trên các kênh và đường dẫn khác nhau.

Cuối cùng, nếu băng thông là tất cả những gì quan trọng nhất, sau đó khi tất cả đã được được đề cập và thực hiện, bất cứ ai có băng thông lớn nhất sẽ thắng. Các nhà cung cấp cáp quang và thiết bị ghép kênh phân chia theo bước sóng và các nhà cung cấp dịch vụ với các đường trục cáp quang lớn có khuynh hướng trở thành những người ủng hộ tích cực nhất của triết lý QoS "chỉ

cần băng thông" (only bandwidth). Điều duy nhất cần quan tâm trên mạng đó là các công ty này sẽ bán cái gì.

Không cần có cuộc tranh luận nào cả. Cả hai phương pháp tiếp cận đều có rất nhiều giá trị. Tuy nhiên, đợi cho đến khi băng thông thực sự không bị ràng buộc, đó là một tương lai không gần, thì việc thêm một vài thông số QoS khác bên cạnh băng thông sẽ là cần thiết và cần phải được quan tâm.

5.3.2 Các vấn đề cơ bản cần giải quyết cho thoại IP và các giải pháp

Các khó khăn lớn nhất mà thoại IP phải đối mặt có thể tóm tắt lại như sau:

- *Trễ* gây ảnh hưởng đến tính tương tác của dịch vụ thoại. Với VoIP, khuyến nghị đưa ra là trễ một chiều nhỏ hơn 150 ms (G.114 của ITU-T).

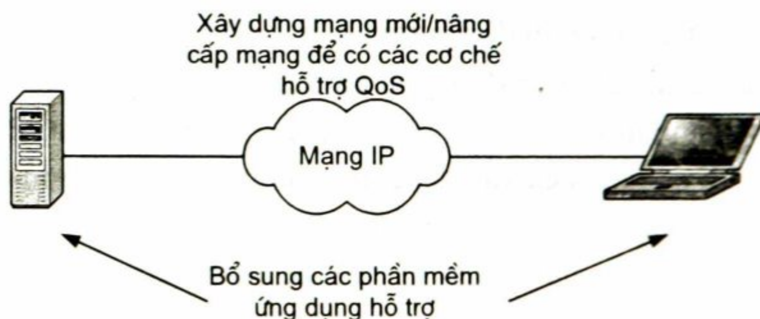
- *Mất gói* ảnh hưởng đến chất lượng cảm nhận của người nghe

- *Biến động trễ* (jitter) gây cản trở đến tính tuần tự và liên tục của luồng thông tin đa phương tiện

Vấn đề nghe có vẻ đơn giản và được giải quyết triệt để nếu mở rộng băng thông thật lớn và như thế sẽ giảm được trễ. Nhưng với nhà khai thác là cả vấn đề, cũng như câu hỏi đặt ra là tại sao người ta không xây dựng hệ thống giao thông thật lớn để không phải lo tắc đường nữa? Đến đây thì ta nhận thấy rằng tiền cho đầu tư xây dựng mạng không nhỏ, ai có thể xây dựng lại cơ sở mạng Internet toàn cầu trong một sớm một chiều? Bài toán kinh tế đặt ra phải khai thác hiệu quả những gì đã có, việc xây dựng mạng với QoS đáp ứng được mọi yêu cầu là việc làm tiếp theo trong tương lai.

Từ những gì đã phân tích ở trên thì các giải pháp nhằm nâng cao chất lượng dịch vụ thoại IP cũng như các dịch vụ đa phương tiện triệt để nhất là xây dựng một mạng với QoS đảm bảo. Các giải pháp như tích hợp các cơ chế để giải quyết vấn đề trễ, biến động trễ và mất gói ở bộ phát và thu sẽ giải quyết được phần nào nhưng tại thời điểm này xem ra lại có vẻ khả thi nhất do đơn giản, đầu tư ít và có thể thực hiện được ngay do không yêu cầu thay đổi cơ sở hạ tầng mạng. Đó chính là những gì chúng ta đã phân tích khi đưa ra các thông số của QoS, giải pháp là các ứng dụng tự bổ sung QoS để đảm bảo chất lượng dịch vụ.

Các giải pháp nhằm nâng cao chất lượng thoại về cơ bản bao gồm 3 cách (hình 5.4): *thứ nhất* nâng cao chất lượng mạng (tức là nâng cao chất lượng cơ sở hạ tầng mạng và xây dựng một mạng đảm bảo về QoS). *Thứ hai là* sử dụng các cơ chế ở đầu thu để làm giảm ảnh hưởng của biến động trễ và khôi phục mất gói. *Thứ ba là* sử dụng cơ chế kết hợp cả đầu thu và phát để sửa lỗi và giảm ảnh hưởng của lỗi.

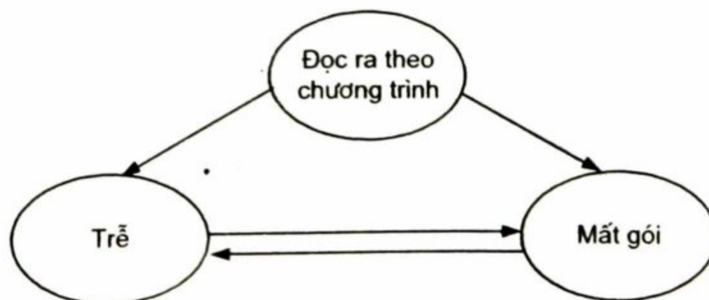


Hình 5.4: Các giải pháp để đảm bảo QoS

Đối với thoại IP, do được nén để hoạt động với tốc thấp nên băng thông cho nó cũng không phải lớn. Khi nói đến độ cảm

nhận về chất lượng thoại, các ảnh hưởng lớn đến chất lượng đó là vấn đề trễ, biến động trễ và mất gói. Còn các thông số QoS khác có thể là thứ yếu. Do đó, về bản chất, những khó khăn cơ bản mà thoại IP phải đối đầu chính là trễ, biến động trễ và mất gói. Tuy nhiên, các thông số này có liên quan mật thiết với nhau.

Ta biết rằng, gói được xác định là mất khi trễ đầu cuối - đầu cuối của gói vượt quá quy định. Như vậy, gói mất là gói không bao giờ đến đích (thời gian trễ bằng vô cùng) nhưng cũng có thể là gói mặc dù đến đích nhưng sau thời điểm quy định đọc ra của chúng (trường hợp này gọi là mất do trễ). Số của gói mất sẽ giảm khi tăng trễ bộ đệm nghĩa là tăng thời gian trễ cho phép của gói. Tuy nhiên, tăng trễ bộ đệm đồng nghĩa với việc tăng trễ đầu cuối - đầu cuối (mà điều này lại ảnh hưởng đến chất lượng, đến độ cảm nhận của người nghe). Do đó, cần phải có sự cân nhắc thỏa đáng giữa hai tham số này (hình 5.5).



Hình 5.5: Mối quan hệ giữa trễ - mất gói và cân bằng giữa chúng

Trong mạng IP, thời gian trễ đối với mỗi gói là khác nhau được gọi là biến động trễ (jitter). Để bù lại biến động trễ đối với mỗi gói khi đến bộ phát, người ta sử dụng giải pháp dùng bộ

đệm thích nghi ở đầu thu trước khi đọc ra. Bộ đệm thích nghi hay bộ đệm đọc ra theo chương trình là bộ đệm được điều khiển bằng chương trình, nó có thể dự đoán được trễ của gói và điều chỉnh thời gian đệm tương ứng với trễ từng gói. Bộ đệm thích nghi sẽ giải quyết được khá tốt vấn đề mất gói do trễ. Tuy nhiên, trễ đầu cuối – đầu cuối vẫn phải đảm bảo theo khuyến nghị G.114 của ITU-T. Trường hợp gói có trễ đầu cuối vượt quá quy định thì coi như bị mất và cần sử dụng các kỹ thuật để tái tạo lại gói. Các kỹ thuật này sẽ được trình bày ở chương sau.

Chương VI

GIẢI PHÁP NÂNG CAO CHẤT LƯỢNG THOẠI

6.1. MẤT GÓI, TRỄ, BIẾN ĐỘNG TRỄ VÀ CÁC GIẢI PHÁP

Nghiên cứu về QoS có thể giúp giải quyết một số vấn đề như: mất gói, biến động trễ và trễ. Nhưng một số vấn đề mà QoS khó giải quyết triệt để được như là trễ mạng, trễ do mã hóa/giải mã, trễ lấy mẫu và trễ do số hóa. Điều quan trọng là phải biết phần nào không thể thay đổi và phần nào có thể điều khiển được theo như bảng sau:

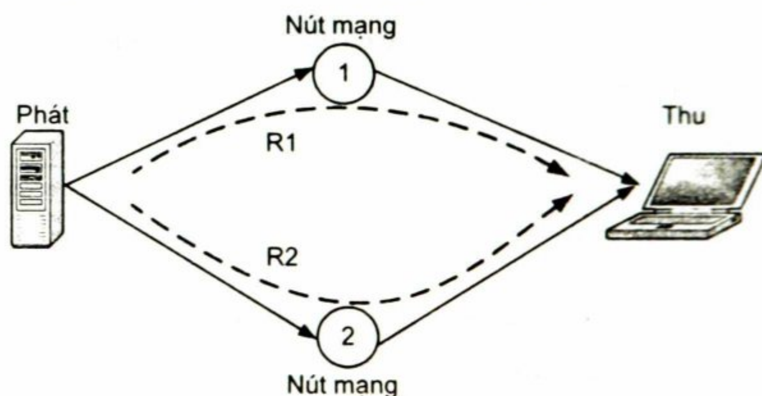
Bảng 6.1: Các thành phần trễ

	Trễ cố định	**Trễ thay đổi**
Trễ xử lý G.729 (5 ms)	5 ms	
Trễ đóng gói G.729 (10 ms/khung)	20 ms	
Trễ mạng		20 - 200 ms (> đến vài giây)
Trễ tại bộ đệm ở đầu thu		2 - 200 ms (> đến vài giây)

Khuyến nghị G.114 của ITU-T cho rằng trễ từ đầu cuối đến đầu cuối không vượt quá 150 ms là duy trì được chất lượng thoại tốt.

Ba yếu tố chính ảnh hưởng đến chất lượng thoại IP mà ta đã phân tích đó là: trễ, biến động trễ và mất gói. Trễ lớn gây ra tiếng vọng và ảnh hưởng đến khả năng tương tác của loại dịch vụ này.

Trong thành phần phân bổ của trễ đầu cuối – đầu cuối gồm có trễ xử lý, trễ mạng và trễ bộ đệm thì trễ mạng là khá lớn và rất biến động. Để khống chế được trễ mạng thì giải pháp hữu hiệu nhất là nâng cao chất lượng mạng và cấu trúc mạng hợp lý. Đây là giải pháp khá tốn kém và phải thực hiện từng bước. Cũng có giải pháp khác không cần nâng cấp mạng nhưng có thể hạn chế bớt trễ và biến động trễ là sử dụng kỹ thuật phát đa đường. Ý tưởng của kỹ thuật này là phía bộ phát sẽ gửi các gói thông tin đồng thời trên các đường khác nhau, tại bộ thu sẽ lựa chọn những gói có cùng nội dung thông tin nhưng đến trước và nội dung thông tin còn toàn vẹn. Như vậy, trễ gói sẽ là MIN[$d_1,d_2,\ldots d_n$], với $d_1,d_2,\ldots d_n$ là trễ của các gói có cùng nội dung đi theo các tuyến R1, R2, ..., Rn như hình 6.1 dưới đây. Và như vậy tỷ lệ mất gói cũng được giảm đáng kể.

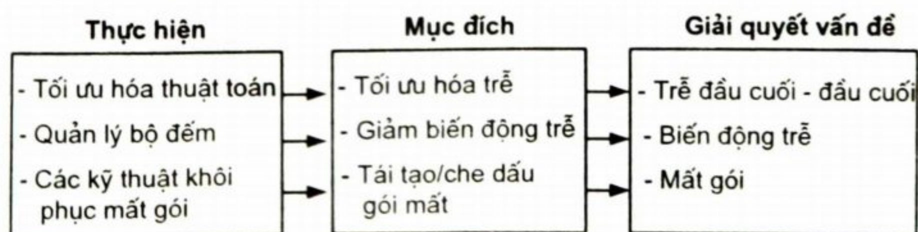


Hình 6.1: Kỹ thuật phát đa đường

Kỹ thuật này chỉ cần thực hiện ở bộ phát và thu, không cần cơ chế hỗ trợ của mạng nhưng có nhược điểm là nó sẽ làm tăng tải của mạng. Kỹ thuật này sẽ rất tốt trong trường hợp thông tin

trao đổi quan trọng nên người ta có thể chấp nhận tăng tải của mạng để đảm bảo chất lượng thông tin. Tuy nhiên, kỹ thuật này sẽ hiệu quả hơn khi kết hợp với một số kỹ thuật khác như khôi phục mất gói, kỹ thuật phát hiện tắc nghẽn để chỉ kích hoạt kỹ thuật phát đa đường khi phát hiện nguy cơ mất gói hàng loạt do nghẽn cục bộ.

Hai thành phần còn lại trong trễ đầu cuối – đầu cuối là trễ xử lý và trễ bộ đệm được giải quyết ở bộ phát và thu. Trễ xử lý bao gồm trễ mã hóa và trễ đóng gói, nó phụ thuộc và thuật toán nén và tốc độ bộ xử lý. Việc tối ưu thuật toán mã hóa sẽ giảm được trễ xử lý. Trễ bộ đệm cũng được giảm thiểu khi sử dụng cơ chế quản lý bộ đệm hợp lý. Và quản lý bộ đệm cũng là mục tiêu giải quyết vấn đề biến động trễ. Điều chỉnh thời gian đọc gói ra khỏi bộ đệm, đồng thời cũng giải quyết tốt vấn đề đối với mất gói. Việc sử dụng kỹ thuật đọc gói thích nghi kết hợp với kỹ thuật khôi phục mất gói sẽ thực hiện tốt mục đích tái tạo/che dấu mất gói. Hình 6.2 dưới đây cho thấy các kỹ thuật sử dụng ở bộ thu phát để giải quyết các vấn đề đối với thoại IP.

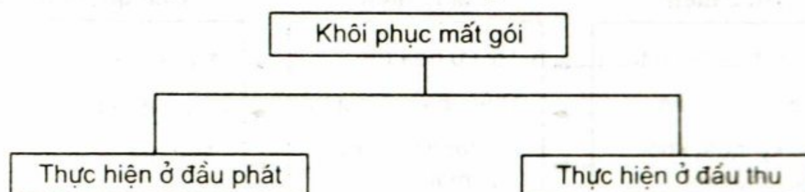


Hình 6.2: Các kỹ thuật thực hiện ở bộ phát và thu

Tái tạo hay che dấu gói mất là kỹ thuật nhằm làm cho người nghe không cảm nhận thấy sự mất mát thông tin. Nó là kỹ thuật nhằm mang lại "âm thanh tự nhiên" nghĩa là các đoạn thông tin

bị mất của cuộc đàm thoại sẽ được "lấp đầy" bởi một đoạn thông tin gần tương tự với đoạn bị mất. Tái tạo là "dựng" lại đoạn thông tin bị mất bằng kỹ thuật nào đó như: sử dụng đoạn thông tin của gói nhận được trước đó hay nội/ngoại suy từ thông tin của các gói trước và sau gói bị mất dựa vào một số đặc tính của thoại như dạng sóng hoặc phổ tần. Che dấu là kỹ thuật mà đúng như tên gọi của nó là thực hiện 'hành động che đậy' những gì bị mất. Việc che dấu tinh vi chính là làm cho người nghe không cảm thấy khó chịu về sự mất thông tin hay độ nghe hiểu của đoạn thoại. Đôi khi tái tạo và che dấu là một sự kết hợp, vì mục đích chung cũng là làm cho thông tin nhận được có chất lượng tốt hơn dựa trên độ cảm nhận của người nghe. Xét trên một nghĩa rộng hơn có thể hiểu là kỹ thuật nhằm khôi phục mất gói tức là làm cho đoạn thoại "trơn mượt" để tai người khó mà cảm nhận thấy sự mất thông tin trong đó như sự gián đoạn, các âm lạ hay sự méo tiếng.

Kỹ thuật khôi phục mất gói có thể được chia thành 2 loại: thực hiện ở đầu phát (sender-based) và thực hiện ở đầu thu (receiver-based) như hình 6.3 dưới đây.



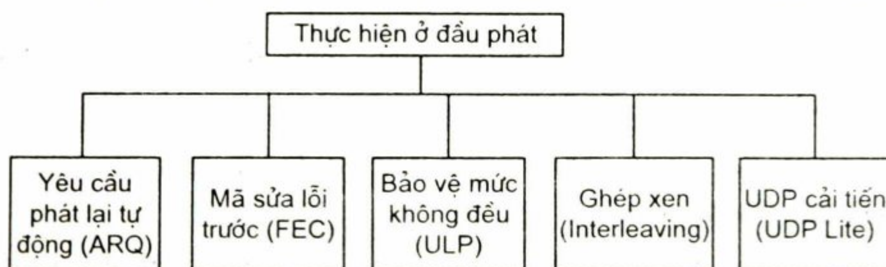
Hình 6.3: Phân loại kỹ thuật khôi phục mất gói

Ngoài giải pháp thực hiện ở đầu phát và đầu thu là các giải pháp thực hiện tại đầu cuối của người sử dụng, có thể áp dụng giải pháp nâng cao chất lượng mạng sử dụng các mô hình hỗ trợ

QoS trong mạng như mô hình dịch vụ tích hợp, mô hình dịch vụ khác biệt, là các mô hình cho mạng trong tương lai. Các mô hình này sẽ được trình bày trong mục 6.4.

6.2. GIẢI PHÁP THỰC HIỆN Ở PHÍA BỘ PHÁT

Cơ chế khôi phục mất gói thực hiện ở đầu phát bao gồm: Yêu cầu phát lại tự động (ARQ - *Automatic Repeat reQuest*), sửa lỗi trước (FEC - *Forward Error Correction*), FEC bảo vệ mức không đều (*Uneven Level Protection*), ghép xen (*Interleaving*) và UDP cải tiến (*UDP Lite*). Hình 6.4 dưới đây phân loại các kỹ thuật khôi phục mất gói thực hiện ở đầu phát.

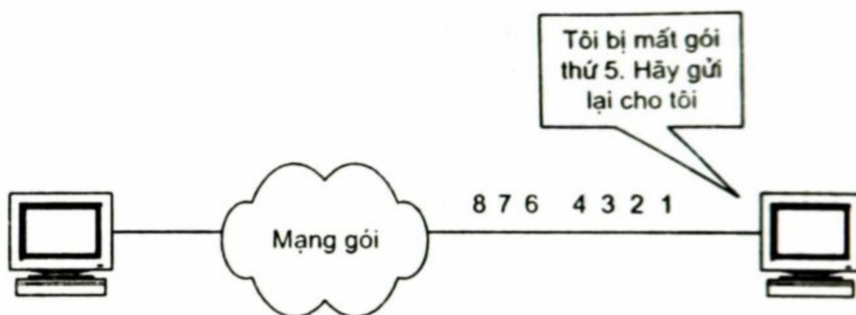


Hình 6.4: Phân loại kỹ thuật khôi phục mất gói thực hiện ở đầu phát

6.2.1. Yêu cầu phát lại tự động (ARQ)

Cơ chế này sẽ thực hiện phát lại các gói lỗi hoặc mất, nó là phương pháp phổ biến để khôi phục dựa trên đầu phát và thường sử dụng trong TCP.

Hình 6.5 cho thấy, phía phát sẽ bổ sung số thứ tự vào các gói được truyền đi là 1, 2, ..., 8. Phía thu có thể căn cứ vào số thứ tự và phát hiện được gói bị mất, ví dụ gói thứ năm như trên hình vẽ.



Hình 6.5: Cơ chế yêu cầu phát lại

Sử dụng ARQ, gói mất được đầu phát phát lại. Cơ chế ARQ bao gồm ba phần:

- Đầu phát hoặc đầu thu phát hiện được dữ liệu mất (timeout).

- Gửi xác nhận: Đầu thu gửi bản tin để xác nhận dữ liệu đã nhận được hoặc bị mất

- Phát lại: Nó sẽ quyết định dữ liệu được đầu phát phát lại.

Cơ chế này đảm bảo truyền dữ tin cậy, chống được mất cụm nhưng nó không thể sử dụng cho các ứng dụng thời gian thực nhạy cảm với trễ như thoại IP bởi vì phát lại gây ra trễ lớn và yêu cầu thêm băng thông.

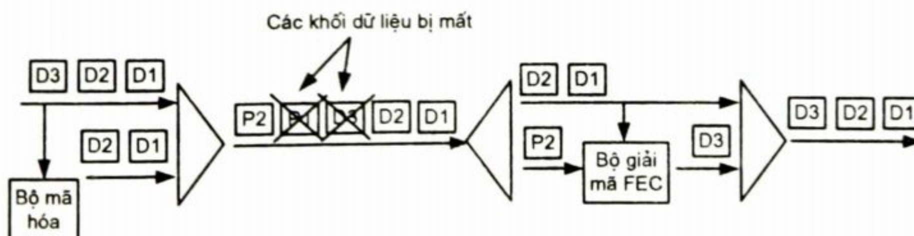
6.2.2. Sử dụng mã sửa lỗi trước (FEC)

Sử dụng mã sửa lỗi FEC, dữ liệu mất sẽ được khôi phục ở đầu thu mà không cần phát lại của đầu phát. Cả dữ liệu gốc và thông tin dư thừa được gửi đến đầu thu. Có hai kiểu thông tin dư thừa là: luồng thông tin độc lập và phụ thuộc.

Thông tin độc lập FEC không cần biết kiểu dữ liệu gốc (thoại, audio, video). Trong thông tin độc lập FEC, dữ liệu gốc cùng với dữ liệu dư thừa đều được gửi tới đầu thu.

Dữ liệu dư thừa nhận được từ dữ liệu gốc nhờ sử dụng toán tử exclusive-OR (XOR) hoặc sử dụng mã Reed-Solomon. Mã Reed-Solomon cho phép bảo vệ lỗi tối ưu nhưng giá thành xử lý cao hơn cơ chế dựa trên toán tử XOR.

Nguyên lý của FEC là truyền đi *k* gói dữ liệu gốc và kèm theo *h* gói chẵn lẻ dư thừa. Ví dụ *k* = 3 và *h* = 2, như vậy các gói truyền đi có 3 gói dữ liệu D1, D2, D3 và hai gói chẵn lẻ P1, P2. Giả sử, một gói dữ liệu D3 và một gói chẵn lẻ P1 bị rớt, thì tại bộ thu có thể khôi phục gói dữ liệu D3 bằng từ các gói dữ liệu nhận được D1, D2 và P2.



Hình 6.6: Cơ chế khôi phục dữ liệu của FEC

FEC làm việc hiệu quả ngay cả khi tỷ lệ *h/k* nhỏ. Ở bộ giải mã FEC, mất gói liên tiếp có thể sửa được với giá trị *k* lớn, vì bộ mã hóa không chỉ sử dụng các bit chẵn lẻ mà còn sử dụng các bit dữ liệu cho việc sửa lỗi. Tuy nhiên, nếu *k* tăng, trễ tái tạo ở bộ thu cũng tăng. Mã FEC có ưu điểm (+) và nhược điểm (-) đó là:

+ Hoạt động của FEC không phụ thuộc vào nội dung của dữ liệu gốc và các gói mất có thể sửa chữa chính xác.

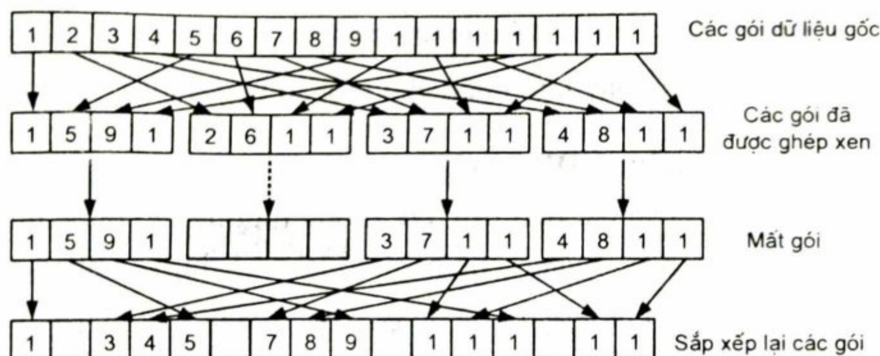
+ Mã FEC yêu cầu trễ cao hoặc yêu cầu thêm băng thông.

6.2.3. Bảo vệ mức không đều (ULP)

ULP được dựa trên thực tế các phần khác nhau của dữ liệu có mức quan trọng không đều nhau được ưu tiên bảo vệ chất lượng khác nhau. Ví dụ, nếu thoại được mã hóa sử dụng CELP thì "pitch" và các tham số bộ lọc dự đoán trước được xem xét là quan trọng hơn các ký hiệu trong khối mã hóa kích thích. Một lỗi trong tham số bộ lọc dự đoán trước có thể dẫn đến giảm chất lượng, trong khi một lỗi trong khối ký hiệu sẽ ảnh hưởng ít thậm trí gần như không ảnh hưởng tới chất lượng cảm nhận thoại. Đây là điểm mấu chốt thúc đẩy sử dụng bảo vệ không đều cho dữ liệu mà có độ quan trọng không đều. Các khối dữ liệu được sắp xếp trong gói RTP giảm dần theo độ quan trọng và như vậy dữ liệu ở phần đầu (quan trọng) sẽ được bảo vệ tốt hơn. Định dạng của tải dữ liệu bao gồm thông tin của các khối dữ liệu được bảo vệ trong gói và các mức bảo vệ, chiều dài bảo vệ của mỗi một mức.

6.2.4. Ghép xen (Interleaving)

Ghép xen là một kỹ thuật khôi phục mất gói được sử dụng cho các ứng dụng mà trễ là vấn đề quan trọng thứ yếu. Nếu cỡ của khối dữ liệu được tạo ra ở một thời điểm bởi một bộ mã hóa mà nhỏ hơn cỡ của tải dữ liệu cho phép trong gói thì một vài khối dữ liệu có thể kết hợp trong một gói. Tuy nhiên, để giảm ảnh hưởng của mất gói (nhất là mất cụm), các khối dữ liệu gốc được ghép xen trước khi phát đi như minh họa trên hình 6.7.



Hình 6.7: Nguyên lý của ghép xen

Tại bộ thu, các khối dữ liệu được tập hợp và sắp xếp lại theo thứ tự như ban đầu và đưa tới bộ giải mã. Ảnh hưởng của mất gói (một gói bao gồm một vài khung dữ liệu) được phân bố trên các khoảng thời gian nhỏ tương ứng thay vì các khối dữ liệu mất liên tiếp. Ảnh hưởng của mất gói giảm bởi vì:

Các khoảng thời gian nhỏ tương ứng có chiều dài nhỏ hơn chiều dài một âm vị (phoneme). Do đó, nếu mất có thể nội suy lại khoảng đã mất và tính dễ hiểu của thoại là không giảm. Trong đó nếu không sử dụng ghép xen thì khi gói bị mất sẽ mất hoàn toàn một âm vị và dẫn tới làm giảm khả năng hiểu được của thoại.

Nếu bộ thu có sử dụng kỹ thuật che dấu mất gói thì việc nội suy có thể thực hiện tốt hơn với khoảng thời gian bị mất ngắn và khó khăn nếu mất trong khoảng thời gian dài.

Ghép xen có một số ưu (+) và nhược (-) điểm là:

+ Giảm ảnh hưởng của mất gói do phân tán mất, biến mất cụm thành mất đơn

+ Không làm tăng tải cho mạng

- Làm tăng trễ gây khó khăn cho một số ứng dụng.

6.2.5. UDP cải tiến (UDP Lite)

UDP Lite là sự cải tiến từ UDP thông thường, nó được thiết kế cho các ứng dụng mà cơ chế hoạt động ở bộ thu vẫn nhận gói dữ liệu bị lỗi chứ không hủy bỏ ngay, việc loại bỏ gói nên để lớp ứng dụng quyết định là dữ liệu còn dùng được hay bỏ. VoIP là ứng dụng thuộc loại này. Lớp IP không giải quyết vấn đề này vì trong tiêu đề IP, trường tổng kiểm tra lỗi (*checksum*) không thực hiện kiểm tra lỗi phần tải dữ liệu trong gói IP. UDP Lite cung cấp một kiểm tra lỗi với mức bao phủ từng phần tùy chọn. Hình 6.8 cho thấy tiêu đề gói UDP đã cải tiến so với tiêu đề UDP thông thường như trong hình 2.3.

Bit 0-3	Bit 4-7	Bit 8-11	Bit 12-15	Bit 16-19	Bit 20-23	Bit 24-27	Bit 28-31
Số cổng nguồn				Số cổng đích			
Thông tin tổng kiểm tra				Tổng kiểm tra lỗi			
Dữ liệu							

Hình 6.8: Định dạng tiêu đề UDP Lite

Sự khác nhau giữa UDP Lite và UDP thông thường là trường chiều dài (length) được đổi thành trường *(checksum coverage)*. Trường *checksum coverage* sẽ tính toán số byte từ byte thứ nhất của tiêu đề UDP. UDP Lite cũng sử dụng một tiêu đề giả bao gồm một vài trường của lớp IP để kiểm tra. Nếu trường *checksum coverage* bằng “0” thì toàn bộ gói UDP dùng để kiểm tra lỗi.

6.2.6. Đánh giá các kỹ thuật thực hiện ở bộ phát

Nếu trễ đầu cuối - đầu cuối (end-to-end) là thứ yếu như trong phát thanh quảng bá thì ARQ, FEC và ghép xen có thể được áp dụng để khôi phục mất gói. Nếu sự tái tạo gói mất có thể chấp nhận thì ghép xen nên được sử dụng vì nó không tăng thêm tải cho mạng. Nếu trễ là vấn đề rất quan trọng thì cơ chế FEC trễ thấp nên được sử dụng. UDP cũng có thể được sử dụng kết hợp với FEC để khai thác sự quan trọng không đều của dữ liệu.

UDP Lite có thể cung cấp ở một lớp ứng dụng với một gói bị lỗi thay vì loại bỏ nó ở một lớp giao thức. Nó lên đến lớp ứng dụng mới quyết định liệu gói lỗi nên sử dụng hay bỏ đi.

Phải thấy rằng qua kỹ thuật được mô tả giúp cho việc lựa chọn phương pháp giải quyết vấn đề khôi phục mất gói, một vài kỹ thuật như ARQ và FEC làm tăng tải mạng nên có thể làm tăng xác suất nghẽn và mất gói. Sự quyết định tốt nhất thường phụ thuộc vào ứng dụng, điều kiện mạng và một số ràng buộc khác.

6.3. GIẢI PHÁP THỰC HIỆN Ở PHÍA BỘ THU

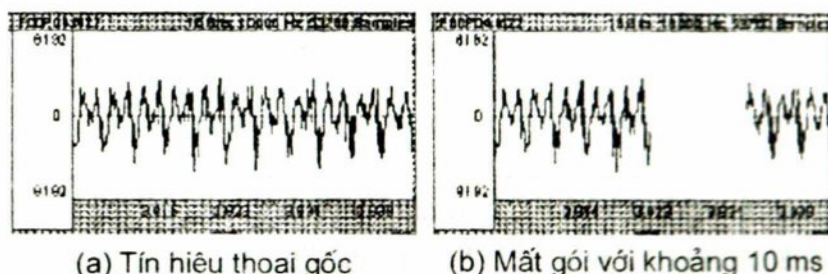
Khôi phục mất gói thực hiện ở đầu thu được xem như bằng cách nào đó để giảm ảnh hưởng đến chất lượng, tác động đến cảm nhận của người nghe, nên việc khôi phục mất gói ở đầu thu có thể xem như việc che dấu sự mất gói này để giảm ảnh hưởng đến chất lượng cảm nhận thấy từ người nghe.

6.3.1. Nguyên lý cơ bản

Trong mục này, chúng ta tổng kết một số kỹ thuật che dấu lỗi mà được thực hiện bởi đầu thu. Các kỹ thuật này sử dụng khi

một cơ chế phía phát sửa hầu hết các mất cụm và chỉ còn một số nhóm nhỏ cần sửa tiếp. Một tỷ lệ mất cụm đã được giảm, che dấu lỗi là một cách hiệu quả để sửa lỗi còn lại.

Phương pháp che dấu mất gói thực hiện thay thế một gói mất. Bằng cách tổng hợp phần âm thanh hoặc tín hiệu thoại tương ứng với gói đã mất. Điều này có thể thực hiện vì hầu hết các phần của tín hiệu thoại hoặc âm thanh thay đổi chậm theo thời gian, do vậy chúng gần như giống các đoạn tín hiệu khác xem hình 6.9.

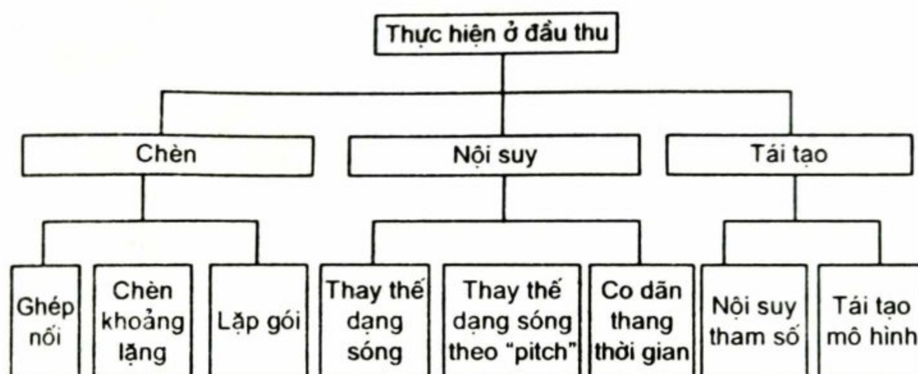

(a) Tín hiệu thoại gốc (b) Mất gói với khoảng 10 ms

Hình 6.9: Các đoạn thoại ngắn có dạng sóng giống nhau

Các kỹ thuật này làm việc tốt khi tỷ lệ mất gói nhỏ (<15%) và khoảng thời gian bị mất nhỏ hơn 40 ms. Khi chiều dài mất liên tiếp gần bằng chiều dài của âm vị thì kỹ thuật này làm việc sẽ không hoàn hảo.

6.3.2. Phân loại

Ở phía thu có 3 phương pháp che dấu mất gói là: Chèn, nội suy và tái tạo. Trong mỗi phương pháp lại chia ra các kỹ thuật thực hiện khác nhau như hình 6.10.



Hình 6.10: Phân loại kỹ thuật khôi phục mất gói thực hiện ở đầu thu

6.3.2.1 Chèn (Insertion)

Trong phương pháp chèn, gói bị mất được thay thế bằng việc chèn đầy vào các bit. Trong kỹ thuật chèn chia ra ba loại như sau:

- Ghép nối (*Splicing*): Đây là kỹ thuật đơn giản nhất, nếu mất gói xảy ra nó sẽ cắt bỏ đoạn tín hiệu tương ứng mất, ví dụ gói thứ 3 bị mất thì tiếp theo gói thứ hai là gói thứ tư. Như vậy trong luồng tín hiệu thoại sẽ có đoạn tín hiệu thoại bị cắt bỏ tương ứng chứa trong gói thứ hai.

- Chèn khoảng lặng (*Silence substitution*): Thay thế chiều dài mất gói với biên độ bằng 0, kỹ thuật này chỉ chấp nhận được khi khung rất ngắn và tỷ lệ mất là rất thấp. Hoặc chèn nhiễu (Noise substitution): Dùng nhiễu nền tương ứng với khoảng lặng thay thế cho đoạn dữ liệu bị mất.

- Lặp lại bằng các gói trước đó: Kỹ thuật này chấp nhận được khi mất gói đơn. Gói tốt nhận được trước gói bị mất được

lặp lại để lấp đầy vào khoảng thoại bị mất. Tuy nhiên, lặp lại gây ra một số âm thanh lạ.

6.3.2.2 Nội suy (Interpolation)

Nội suy là kỹ thuật phức tạp hơn chèn, vì chúng còn xem xét tới đặc tính thay đổi của tín hiệu thoại hoặc âm thanh. Phương pháp này sẽ sử dụng các gói lân cận (trước và sau gói bị mất) để làm cơ sở tái tạo lại gói bị mất và thay thế chúng. Kỹ thuật nội suy có 3 dạng:

- Thay thế bằng dạng sóng (*waveform substitution*) của tín hiệu thoại hoặc âm thanh trước hoặc sau đoạn bị mất.

- Nội suy dạng sóng theo chu kỳ "pitch" (*Pitch waveform interpolation*): Nếu gói mất là đoạn vô thanh được thay thế bằng cách lặp lại. Nếu gói mất là đoạn hữu thanh được thay thế bằng cách lặp lại dạng sóng theo chu kỳ "pitch". Kỹ thuật này cho chất lượng tốt hơn "waveform substitution".

- Co dãn thang thời gian (*Time scale modification*): Kỹ thuật này cho phép co dãn đoạn thoại hay audio trước và sau đoạn bị mất để bù lấp những đoạn bị mất này. Đoạn sóng chuyển từ gói trước sang gói sau được nội suy trong một cửa sổ để tìm dạng sóng tối ưu, do đó chất lượng của kỹ thuật này cao hơn các kỹ thuật đã trình bày ở trên. Tuy nhiên, trễ thuật toán cũng cao hơn và tính toán phức tạp hơn.

6.3.2.3 Kỹ thuật tái tạo (Regeneration repair techniques)

Kỹ thuật tái tạo sử dụng những điều đã biết của thuật toán mã hóa để tổng hợp tín hiệu thoại hoặc âm thanh tương ứng với gói

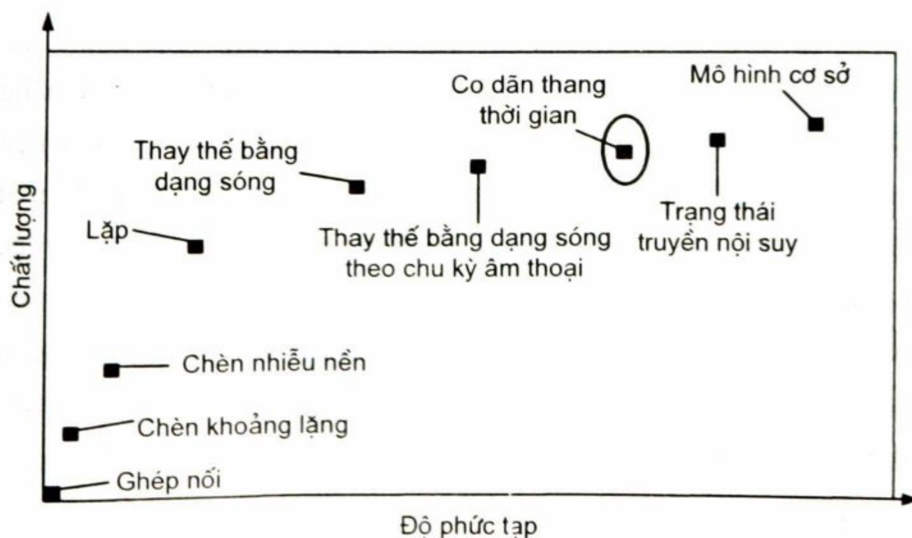
đã mất, công nghệ này phụ thuộc vào bộ mã hóa thoại nhưng chất lượng tốt bởi vì số lượng lớn trạng thái thông tin được sử dụng để tái tạo. Tuy nhiên, kỹ thuật này cũng yêu cầu tính toán khá phức tạp và trễ tính toán cũng lớn. Kỹ thuật này chia làm 2 loại:

- Nội suy tham số (*Parameter interpolation*): Các tham số mã hóa như LPC và độ tăng ích có thể được sử dụng để nội suy.

- Khôi phục trên cơ sở mô hình (*Model-based recovery*): Đoạn mất của tín hiệu được khôi phục bằng cách mô hình hóa dạng mã hóa của tín hiệu thoại hoặc âm thanh của nó.

6.3.3 Đánh giá các kỹ thuật thực hiện ở bộ thu

Hình 6.11 dưới đây cho thấy sự liên quan giữa độ phức tạp và chất lượng của các phương pháp khôi phục mất gói khác nhau.



Hình 6.11: Mối quan hệ giữa chất lượng và độ phức tạp của các phương pháp khôi phục mất gói

Trong các kỹ thuật trên thì kỹ thuật chèn là đơn giản nhưng chất lượng không cao. Các kỹ thuật nội suy và tái tạo cho chất lượng khá tốt nhưng độ phức tạp cũng khá cao.

Sau đây sẽ trình bày chi tiết kỹ thuật có chất lượng khá cao và dễ thực hiện đó là kỹ thuật co dãn thang thời gian. Kỹ thuật co dãn thang thời gian được sử dụng kết hợp với kỹ thuật đọc ra thích nghi sẽ đạt được hiệu quả cao.

6.3.4. Cơ chế đọc ra thích nghi và kỹ thuật co dãn thang thời gian

Trong hệ thống thông tin, để giảm trễ, giảm biến động trễ và khôi phục mất gói có thể được thực hiện với sự kết hợp của các lớp khác nhau trong mô hình OSI. Một chức năng quan trọng được thực hiện ở đầu thu là che dấu mất gói, nghĩa của che dấu ở đây thể hiện khả năng "lấp đầy" những đoạn thoại bị mất bằng kỹ thuật co dãn về mặt thời gian, mục đích che dấu sự mất mát làm người nghe không cảm nhận thấy sự mất thông tin do mất gói gây nên). Một chức năng khác được đưa ra và cũng rất hiệu quả là sử dụng cơ chế đọc ra thích nghi để giảm mất các gói đến trễ (nếu sử dụng cơ chế đọc ra thông thường thì các gói này sẽ bị mất). Cơ chế này có thể hiểu là tính toán thời điểm đọc từng gói ra thay đổi thích ứng với biến động trễ của từng gói.

Ý tưởng của kỹ thuật co dãn thang thời gian (TSM – *Time Scale Modification*) của tín hiệu thoại được sử dụng không phải để thay đổi toàn bộ tốc độ đọc ra của luồng tín hiệu thoại mà là để sắp xếp lại (cấu trúc lại) đoạn tín hiệu bị mất hoặc trễ. Ý tưởng chính

là để tạo ra đoạn tín hiệu có chu kỳ âm lượng tiếng nói (pitch) phù hợp với tín hiệu gốc, bằng cách sử dụng thông tin chu kỳ giọng nói của các khung trước và sau khung bị mất. Kỹ thuật TSM sử dụng thuật toán "lấp đầy bằng sóng đồng dạng" (WSOLA - Wave Similarity Overlap-Add) có thể tạo ra tín hiệu với cùng chu kỳ giọng nói từ tín hiệu được cung cấp cho thuật toán.

6.3.4.1 Cơ chế đệm và chương trình đọc ra

Trong mạng "cố gắng tối đa" như mạng Internet, các gói thoại hay audio có thể đến trễ hoặc thậm chí mất do không đến được đích vì lý do nào đó. Ở đầu phát, các gói thoại/audio được tạo ra trong những khoảng thời gian đều nhau và được gửi qua mạng IP đến đầu thu. Các gói có thể đi theo nhiều đường để đến đích và thời gian đến phụ thuộc vào trạng thái và khoảng cách đường đi, do đó trễ mạng của các gói khác nhau là điều dễ hiểu, nó được hiểu là biến động trễ (jitter). Các phương pháp để giảm biến động trễ có thể chia làm ba phương pháp như sau:

- Phương pháp thực hiện phía phát: Phía phát sẽ lựa chọn chế độ mã hóa và điều chỉnh tốc độ bít theo trạng thái mạng hiện thời. Tổng trễ đầu cuối, biến động trễ và tỷ lệ mất gói đo bằng mức độ nghẽn của mạng.

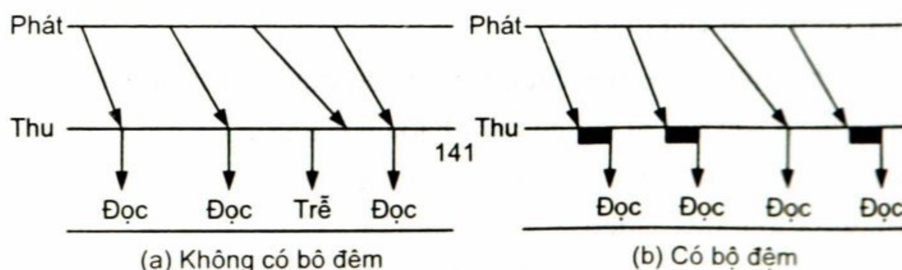
- Phương pháp thực hiện trên cơ sở mạng: Tài nguyên được cấp phát ở các nút mạng để đảm bảo chất lượng dịch vụ yêu cầu. Như vậy, các nút mạng có cơ chế nhận biết sự ưu tiên của các gói như thoại hay video.

- Phương pháp thực hiện phía thu: Các gói được đệm trong một khoảng thời gian ngắn trong bộ đệm trước khi được đọc ra.

Các gói đến sau thời hạn đọc ra coi như bị mất. Bằng cách tăng trễ bộ đệm thì số gói bị mất sẽ giảm nhưng lại làm tăng trễ đầu cuối - đầu cuối. Do đó cần có sự cân nhắc giữa tỷ lệ mất gói và trễ gói đầu cuối - đầu cuối trung bình.

Trong ba phương pháp này thì phương pháp thứ nhất thực hiện phía phát yêu cầu thay đổi cấu trúc của cả bộ mã hóa và giải mã. Phương pháp thứ hai yêu cầu phải thay đổi cấu trúc cơ sở hạ tầng mạng. Phương pháp thứ ba thực hiện ở phía thu là có vẻ hấp dẫn khi chúng chỉ yêu cầu thêm bộ đệm đọc ra ở đầu thu mà có thể giải quyết vấn đề khá hiệu quả.

Trong hình 6.12 mô tả hai trường hợp không có bộ đệm và có bộ đệm. Trường hợp không có bộ đệm, các gói được đọc ra ngay khi nhận được. Hình 6.12(a) cho thấy khi gói đầu tiên đến nó sẽ được thiết lập thời gian đọc ra, các gói tiếp theo sẽ được đọc ra theo một chu kỳ trễ liên quan với gói đầu tiên. Tuy nhiên, những gói trải qua trễ mạng lớn sẽ đến quá muộn. Trong trường hợp này, tỷ lệ đáng kể các gói bị mất và chất lượng thoại là rất thấp khó chấp nhận. Bộ đệm các gói với chu kỳ ngắn trên hình 6.12(b) đảm bảo rằng sẽ có ít gói bị mất hơn.



Hình 6.12: Vai trò của bộ đệm

Phương pháp chung để điều khiển đọc ra các gói được thực hiện bằng bộ đệm đọc ra ở bộ thu để hấp thu biến động trễ trước khi luồng thoại đọc ra. Khi sử dụng kỹ thuật hấp thu biến động trễ, các gói không đọc ra ngay lập tức khi nhận được mà giữ chúng trong bộ đệm và được đọc ra theo thời gian được lập trình. Kỹ thuật này sẽ tăng thêm trễ cho các gói đến sớm nhưng nó cho phép sử dụng được các gói đến có độ trễ cao thay vì các gói này coi như bị mất nếu không sử dụng cơ chế đệm đọc ra. Lưu ý rằng có sự cân bằng (thoả hiệp) giữa thời gian trung bình mà các gói lưu trong bộ đệm (trễ bộ đệm) và số các gói bị mất (do đến quá muộn).

Thời hạn cuối cùng (*deadline*) cho gói đến trễ mà tăng thì có thể đọc ra nhiều gói hơn, nghĩa là tỷ lệ mất thấp nhưng đồng thời giá phải trả cho trễ bộ đệm cũng cao hơn. Trễ bộ đệm góp phần đáng kể làm tăng trễ đầu cuối - đầu cuối, mà điều này yêu cầu rất chặt chẽ để không ảnh hưởng đến tính chất dịch vụ (nhất là với các dịch vụ thời gian thực như thoại). Tuy nhiên, việc tuân thủ theo khuyến nghị của ITU-T G.114 là trễ một chiều dưới 150 ms cho cả hai hướng đối với thông tin thoại là có thể thực hiện được. Trong thực tế, mất gói trong các ứng dụng nhạy cảm với trễ như VoIP thì không chỉ các gói bị rớt do nghẽn mạng mà còn do biến động trễ. Ảnh hưởng của biến động trễ và cơ chế của thực hiện đệm có tác động đáng kể đến chất lượng của ứng dụng. Tuy đệm đọc ra giảm được một số gói trễ ngắn, còn một số gói có trễ đầu cuối - đầu cuối lớn thì trễ đọc ra sẽ vẫn được coi như là mất, tuy

nhiên nếu tỷ lệ mất gói dưới 20% có thể áp dụng các phương pháp che dấu mất gói (PLC - Packet Loss Concealment). Sử dụng thuật toán đọc thích nghi là giải pháp điều chỉnh trễ đọc ra thay đổi tuỳ theo điều kiện biến động của mạng.

Các nghiên cứu trước đây tập trung vào việc cân bằng giữa trễ và mất gói, trong khi cố gắng bù hoàn toàn hoặc gần như hoàn toàn biến động trễ của các "talkspurt". Bằng cách thiết lập thời gian cố định cho tất cả các gói trong một "talkspurt". các gói ra được đọc như tốc độ gốc, liên tục và theo chu kỳ mẫu. ví dụ các khoảng 20 ms. Tuy nhiên, trong các đoạn thoại có "talkspurt" dài và có sự biến động lớn của các gói trong "talkspurt" thì phương pháp này chưa đáp ứng được. Do vậy mà cần phải có phương pháp đọc thích nghi không chỉ với các "talkspurt" mà trong từng "talkspurt", cụ thể là thích nghi với biến động trễ của từng gói.

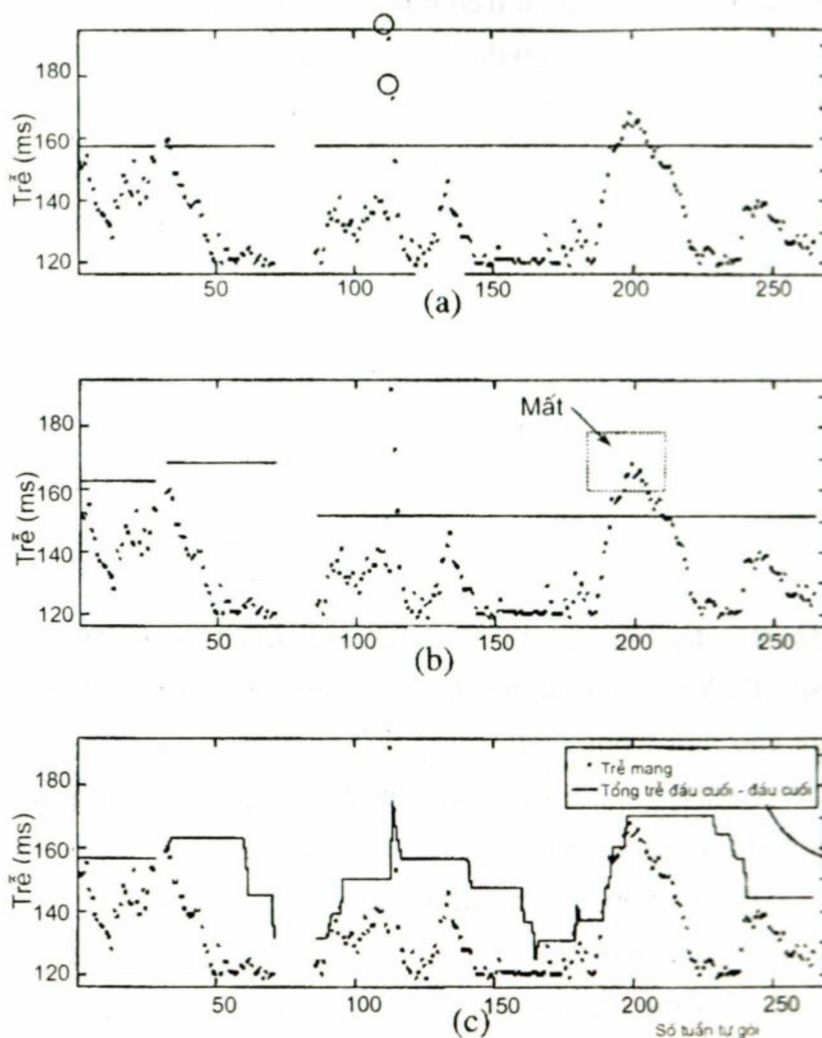
6.3.4.2 Phương pháp đọc ra cố định và thích nghi

Như trên đã trình bày, để giảm biến động trễ có 3 phương pháp, trong đó phương pháp thứ ba là chỉ thực hiện ở phía thu dễ thực hiện hơn cả và cũng rất hiệu quả. Sau đây sẽ trình bày các cơ chế sử dụng bộ đệm và cơ chế đọc từ bộ đệm ra thích hợp để giải quyết vấn đề biến động trễ và mất do trễ.

Hình 6.13 minh họa ba cơ chế thiết lập chương trình đọc cơ bản. Hình vẽ cho thấy trễ của các gói thoại trên mạng như đường chấm chấm và trễ đọc ra như đường nét liền. Các gói đến sau thời hạn cho phép (nằm trên đường nét liền) sẽ coi như là mất.

Cơ chế đơn giản nhất trên hình 6.13(a) là thuật toán 1, nó sử dụng thời hạn đọc ra là cố định cho tất cả các gói thoại trong một phiên (hay một cuộc gọi thoại).

Với các thuật toán 2 hình 6.13(b), cơ chế đọc ra đã được cải tiến so với thuật toán 1, trễ mạng được giám sát và thời gian đọc ra được điều chỉnh thích nghi sau mỗi một chu kỳ của khoảng lặng (nghĩa là trong một "talkspurt" trễ đọc ra là không thay đổi). Dựa trên sự quan sát cho một cuộc thoại điển hình cho thấy, đoạn tín hiệu thoại là tập hợp của các "talkspurt" nối với nhau bởi các khoảng lặng. Thời gian đọc ra của mỗi "talkspurt" mới có thể được điều chỉnh bằng cách dãn hoặc nén các chu kỳ khoảng lặng. Tuy nhiên, thuật toán này bị hạn chế khi các "talkspurt" dài và sự biến động trễ mạng cao trong các "talkspurt". Ví dụ nó không thể thích nghi với trường hợp đỉnh của trễ cao như trong "talkspurt" thứ ba ở các gói được khoanh tròn trên hình 6.13. Như các kết quả đã nghiên cứu cho thấy, các gói bị mất theo cụm sẽ làm giảm khả năng nghe rõ. Mỗi một gói có thể có thời gian đọc ra được lập trình khác nhau tùy theo thống kê trễ biến đổi. Phương pháp này được biểu diễn bằng thuật toán 3 trong hình 6.13(c). Thuật toán 3 yêu cầu co dãn các gói thoại để bảo đảm đọc ra liên tục và do vậy cần phải co dãn thời gian đọc ra lớn. Tuy nhiên, sự mềm dẻo cho phép giảm trễ bộ đệm trung bình trong khi vẫn giảm được mất gói do đến trễ ở cùng thời điểm. Nhờ vậy mà sự cân bằng giữa trễ bộ đệm và mất do trễ được cải thiện tốt hơn.

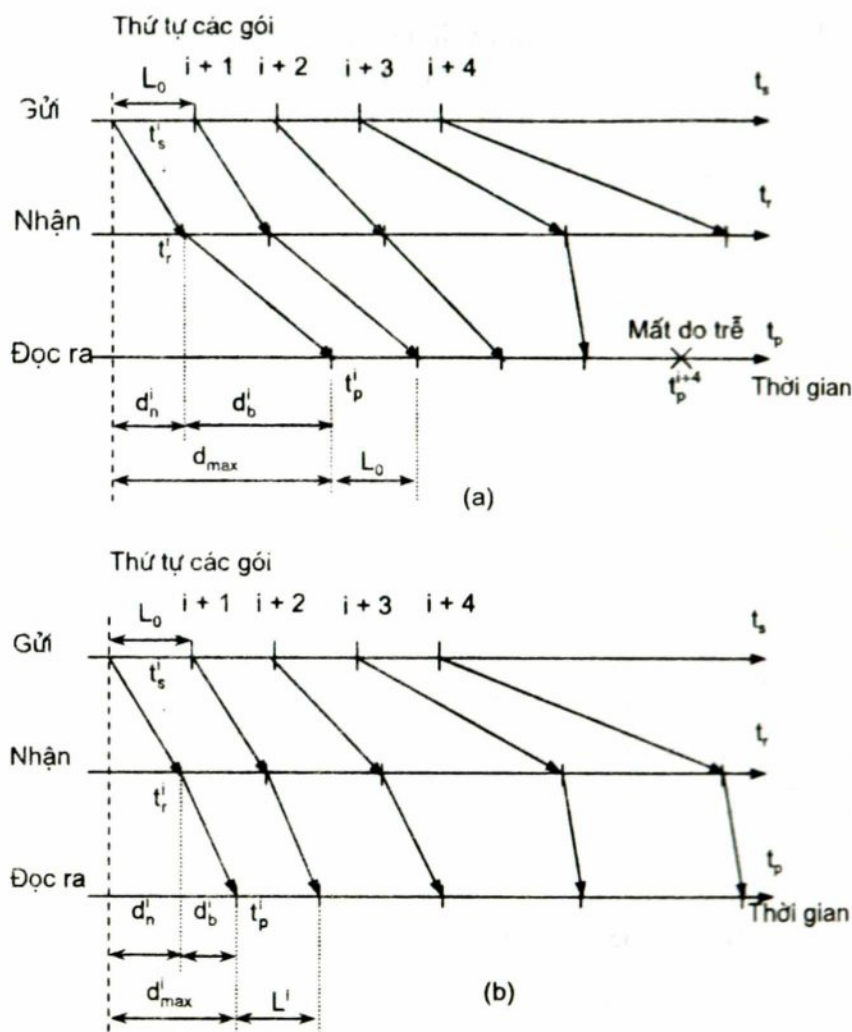


Hình 6.13: Các cơ chế đọc ra khác nhau. (a) thuật toán 1: cơ chế đọc ra cố định. (b) thuật toán 2: cơ chế đọc ra điều chỉnh giữa các talkspurt. (c) thuật toán 3: cơ chế đọc ra thích nghi điều chỉnh trong từng talkspurt

Bảng 6.2 định nghĩa một tiêu chuẩn đánh giá chất lượng cho các cơ chế thiết lập thời gian đọc. Ký hiệu cơ bản và định nghĩa một số tham số như trễ đệm trung bình, tỷ lệ mất do trễ như là tiêu chuẩn chất lượng cơ bản để thuận tiện cho việc tính toán.

Bảng 6.2: Các ký hiệu cơ bản để tính toán thời gian đọc ra

Ký hiệu	**Mô tả**
t_s^i	Thời điểm gói *i* được gửi
t_r^i	Thời điểm nhận được gói *i*
t_p^i	Thời điểm gói *i* được đọc ra
d_n^i	Trễ mạng của gói *i*
d_b^i	Trễ bộ đệm của gói *i*
d_b	Trễ đệm trung bình của một luồng
d_t^i	Tổng trễ của gói *i*
d_{max}	Thời hạn đọc ra cố định
d_{max}^i	Thời hạn đọc ra của gói *i*
D^k	Thống kê bậc đã sắp xếp của $\{d_n^i\}$
ε_n	Tỷ lệ mất liên kết
ε_l	Tỷ lệ mất do trễ
$\hat{\varepsilon}_l$	Tỷ lệ mất do trễ xác định bởi người sử dụng
ε_b	Tỷ lệ mất cụm
ε	Tỷ lệ mất tổng
R	Tập hợp các gói nhận được
P	Tập hợp các gói được đọc
B	Tập hợp các gói mất liên tiếp
N	Số các gói trong luồng
L_0	Thời gian đóng gói ở phía gửi
Li	Chiều dài thực tế của gói i đã co dãn
$\hat{L}^i$	Chiều dài mong muốn của gói i



Hình 6.14: Đọc ra cố định (a) và đọc ra thích nghi (b)

Như minh họa trong hình 6.14 dưới đây, biểu diễn thời gian khi một gói được gửi, nhận và đọc ra là t_s^i, t_r^i, t_p^i. Chỉ số i = 1, 2,... N biểu thị số thứ tự các gói, giả sử có N gói được gửi trong đoạn thoại. Vì các gói thoại thường được xử lý và đóng gói trong

các khối cố định và các gói được gửi đi theo chu kỳ với một thời gian đóng gói bằng hằng số L_0, ví dụ $t_s^{i+1} - t_s^i = L_0 = const$. Trễ bộ đệm của gói i được tính bởi $d_b^i = t_p^i - t_r^i$, trong khi trễ mạng d_n^i được tính bởi $d_n^i = t_r^i - t_s^i$. Trễ tổng d_t^i là tổng của hai trễ trên $d_t^i = d_n^i + d_b^i$. Chú ý rằng trễ tổng không bao gồm thời gian mã hóa và đóng gói (là một trễ thành phần và bằng hằng số của trễ đầu cuối - đầu cuối), vì mối quan tâm chính là truyền gói và đọc ra. Chúng ta sử dụng $d_n^i = \infty$ để chỉ ra rằng gói thứ i bị mất trong khi truyền và không bao giờ có thể đến đích được. Do vậy, tập hợp các gói nhận được tính bởi $R = \{i | t_r^i < \infty\}$.

Nhiệm vụ của cơ chế đọc ra cụ thể là thiết lập trễ tổng cho phép tối đa có thể d_{max}^i (thời hạn đọc ra) cho mỗi gói. Chú ý rằng, trong thuật toán 1 và 2 trong hình 6.13 thì $d_{max}^i = d_{max}$ = hằng số cho tất cả các gói i thuộc vào phiên hoặc cùng "talkspurt". Do đó, thời gian đọc ra các gói ở bộ thu cố định như minh họa trên hình 6.13(a) và chiều dài của các gói ra là hằng số, ví dụ $L_0 = t_p^{i+1} - t_p^i$. Với thuật toán 3 sự thích ứng được thực hiện thực sự trên từng gói. Như kết quả cho thấy, chiều dài của thoại được đọc ra có thể khác đối với mỗi gói.

$$L^i = t_p^{i+1} - t_p^i \qquad (6.1)$$

Ở đây L_i là chiều dài thực sự của gói i sau khi đã dãn. Kết quả đọc ra thích nghi được minh họa trong hình 6.13(b). Co dãn thời gian đọc được yêu cầu cho từng gói thoại là mục tiêu của

của kỹ thuật này. Khi đánh giá các cơ chế đọc thích nghi khác nhau, chúng ta chú ý đến hai tham số, thứ nhất là trễ đệm trung bình được tính bởi.

$$d_b = \frac{1}{|P|}\sum_{i\in P} d^i_{max} - d^i_n \qquad (6.2)$$

Trong đó $P = \{i|t^i_p > t^i_r\}$ là tập hợp của các gói được đọc ra và $|P|$ biểu thị các yếu tố của tập hợp này. Tham số thứ hai là tỷ lệ mất do trễ kết hợp, được tính bởi:

$$\varepsilon_\ell = (|\Re| - |P|)/N \qquad (6.3)$$

Để đầy đủ chúng ta cũng định nghĩa tỷ lệ mất liên kết $\varepsilon_n = (N - |\Re|)/N$. Tỷ lệ mất tổng bằng tổng của hai tham số $\varepsilon = \varepsilon_n + \varepsilon_\ell$. Và sau cùng là tỷ lệ mất cụm ký hiệu bởi ε_b, để xác định số lượng cụm bị mất. Mất cụm được xem xét riêng bởi vì chúng khó che dấu và làm giảm đáng kể chất lượng thoại. Định nghĩa tập hợp các gói với trường hợp mất hai gói liên tiếp là $B = \{i|t^i_p < t^i_r, t^{i+1}_p < t^{i+1}_r\}$, khi đó tỷ lệ mất cụm được tính bởi $\varepsilon_b = |B|/N$. Mặt khác, tỷ lệ mất cụm ε_b là phần trăm xảy ra mất cụm trong một phiên.

Trong công thức tính ε_b, M các gói bị mất liên tiếp được tính như M - 1, trong đó $M \geq 2$.

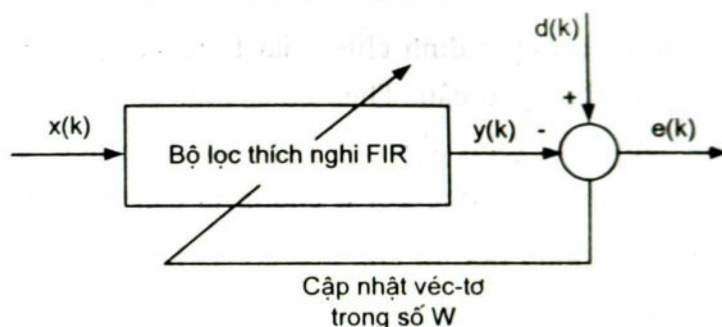
6.3.4.3 Tính toán thời gian đọc ra

Hoạt động cơ bản của chương trình đọc ra là để thiết lập thời gian t_p cho mỗi gói. Trước khi gói i có thể được đọc ra

chúng ta cần phải quyết định chiều dài L_i để thực hiện xác lập thang thời gian theo yêu cầu. Theo công thức (6.1), có thể hoàn toàn thiết lập thời gian đọc ra của các gói tiếp theo với $t_p^{i+1} = t_p^i + L^i$. Do đó, để đọc gói i, chúng ta cần ước lượng thời gian đến và thời gian đọc ra của gói i + 1, hoặc ước lượng trễ mạng d_n^{i+1}. Nếu trễ gói tiếp theo được ước lượng đúng tức là nó sẽ đến đúng thời hạn và sẽ sẵn sàng khi gói hiện tại vừa được đọc ra xong. Việc ước lượng tốt trễ mạng là yếu tố quan trọng cho chương trình đọc ra thích nghi. Có khá nhiều thuật toán để tính toán thời gian trễ đọc ra như: thuật toán dựa trễ ước lượng tự hồi quy, thuật toán dựa trên thống kê, thuật toán dựa trên lọc thích nghi. Trong đó thuật toán dựa trên lọc thích nghi có nhiều ưu điểm hơn và đáp ứng được sự biến đổi nhanh của trễ mạng.

Trong phương pháp sử dụng thuật toán đệm đọc ra thích nghi, thay vì phản ứng lại sự thay đổi bất thường của mạng, nó sẽ dự đoán trễ mạng. Trễ đọc ra được tính từ trễ và biến động trễ mạng đã dự đoán. Dự đoán chính xác trễ mạng có thể đáp ứng được sự thay đổi nhanh của mạng do điều chỉnh trễ hiệu quả hơn.

Thuật toán lọc thích nghi truyền thống đã được sử dụng cho việc làm bằng và triệt tiếng vọng. Thuật toán lọc thích nghi cơ sở mục đích là tối thiểu hóa sai số bình phương trung bình mong đợi giữa giá trị thực và giá trị ước lượng. Các số liệu trước đó được cho qua bộ lọc FIR để tính toán ước lượng hiện thời. Sai số bình phương trung bình sau đó được sử dụng để hiệu chỉnh trọng số của bộ lọc thích nghi. Sơ đồ khối của bộ lọc thích được minh họa trên hình 6.15.



Hình 6.15: Sơ đồ khối của bộ lọc LMS thích nghi

Thuật toán bình phương trung bình nhỏ nhất tiêu chuẩn hóa (NLMS) được sử dụng cho bộ ước lượng thích nghi. Ước lượng trễ mạng $\hat{d}_i$ được tính như sau:

$$\hat{d}_i = w_i^T n_i \tag{6.4}$$

Trong đó $\hat{d}_i$ là giá trị trễ mạng đã dự đoán cho gói thứ i, W_i là véc tơ Nx1 của các hệ số lọc thích nghi, w^T là chuyển vị của w và n_i là véctơ Nx1 của trễ mạng của N gói trước đó. Trọng số của bộ lọc w_i được cập nhật sau mỗi gói sử dụng thuật toán NLMS như công thức (6.5):

$$w_{i+1} = w_i + \frac{\mu}{n_i^T n_i + a} \cdot n_i . e_i \tag{6.5}$$

Trong đó μ là độ lớn của bước nhảy, a là một hằng số nhỏ được thêm vào để tránh chia cho 0 và e_i là sai số ước lượng. Sai số ước lượng được tính như sau:

$$e_i = \hat{d}_i - n_i \tag{6.6}$$

Trong đó n_i là trễ mạng thực tế và $\hat{d}_i$ là trễ dự đoán của gói i. Bảng 6.3 liệt kê các giá trị của của các tham số sử dụng.

Bảng 6.3: Các giá trị tham số được sử dụng trong thuật toán NLMS

Tham số	Giá trị
w_0	$[1\ 0 \ldots 0]^T$
N	18
μ	0,01

Biến động trễ mạng $\hat{v}_i$ và trễ đầu cuối - đầu cuối tổng D_i được tính theo công thức (6.7) và (6.8) như được mô tả cho giao thức điều khiển truyền trong RFC 793.

$$\hat{v}_i = \alpha \hat{v}_{i-1} + (1-\alpha)\,|\,\hat{d}_i - n_i\,| \quad (6.7)$$

$$D_i = \hat{d}_i + \beta \hat{v}_i \quad (6.8)$$

Trong đó α là hệ số điều khiển tốc độ hội tụ của thuật toán, β là hệ số an toàn được sử dụng để làm nhẹ bớt sự cân bằng giữa trễ đầu cuối - đầu cuối và mất do trễ. Giới hạn $\beta \hat{v}_i$ là giới hạn bộ đệm an toàn được sử dụng để bảo đảm là trễ đầu cuối - đầu cuối đủ lớn để chỉ một phần nhỏ của các gói nhận được sẽ đến quá muộn không được đọc ra. Giá trị β cao dẫn đến tỷ lệ mất thấp hơn, tuy nhiên, trễ đầu cuối - đầu cuối lại tăng.

Thuật toán WSOLA co dãn gói cho đọc ra thích nghi

Như mô tả ở trên, đọc ra thích nghi chỉ đạt được khi từng gói thoại riêng có thể được xác định tỷ lệ co dãn mà không làm giảm chất lượng thoại.

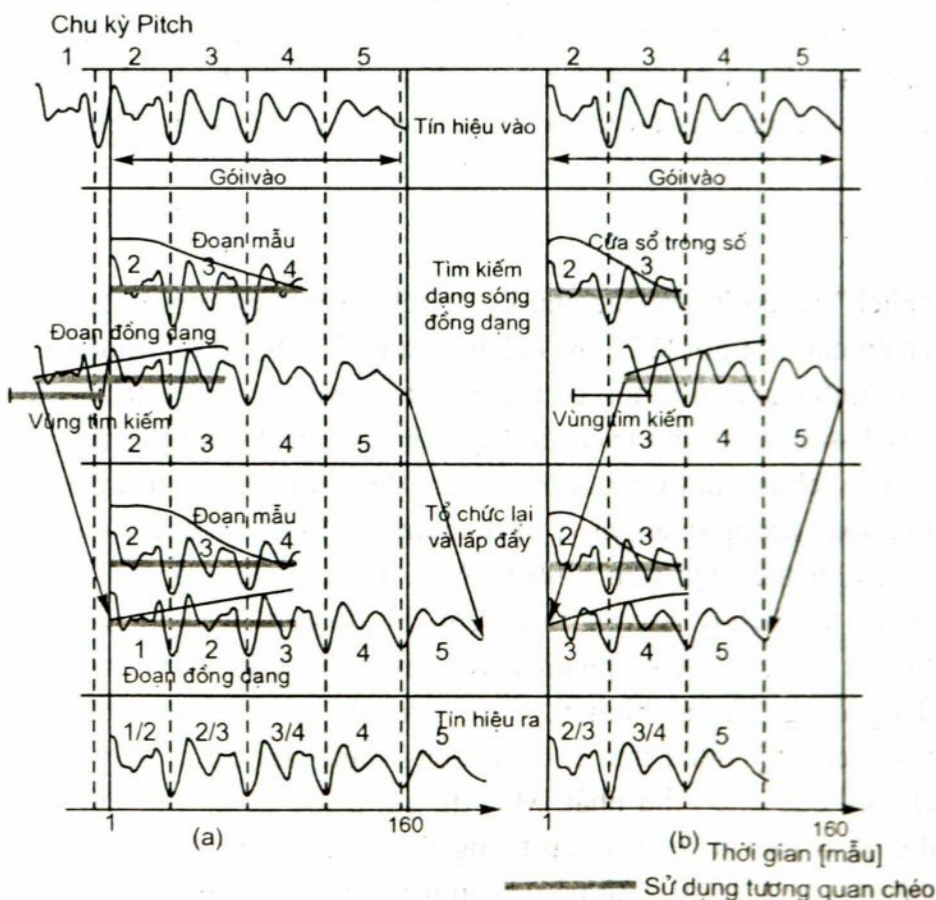
Thuật toán "lấp đầy bằng sóng đồng dạng" (WSOLA - Wave Similarity Overlap-Add) được sử dụng cho kỹ thuật TSM của thoại. Kỹ thuật TSM mục đích là thay đổi tốc độ nói từng phần của thoại nhưng vẫn giữ được độ cảm nhận của thoại như âm sắc, chất lượng thoại và chu kỳ thay đổi biên độ âm lượng (chu kỳ "pitch").

Sự xác lập tỷ lệ co dãn gói thoại được nhận biết bởi kỹ thuật co dãn thang thời gian dựa trên thuật toán WSOLA, nó là một phương pháp trên cơ sở nội suy hoạt động trong toàn bộ miền thời gian. Kỹ thuật này đã được sử dụng để thiết lập tỷ lệ các khối audio dài, sửa và cải tiến để che dấu mất gói bằng cách dãn khối của một vài gói. Ý tưởng cơ bản của WSOLA là phân tích đầu vào thành các đoạn với chiều dài tương đương và các phần đầu chồng lấp lên nhau. Đặc biệt, WSOLA tạo ra dạng sóng tổng hợp y(n), mà vẫn bảo đảm sự đồng dạng tối đa với dạng sóng gốc x(m). Sự sắp xếp lại dẫn đến chiều dài thoại ra được thay đổi. Để các đoạn này được bổ sung trong phần chồng lấp, các vị trí liên quan của chúng trong đầu vào được tìm thấy qua tìm kiếm tương quan chéo tối đa giữa chúng, để chúng có sự đồng dạng tối đa và sự chồng lấp sẽ không gây nên bất cứ sự gián đoạn gì ở đầu ra. Cửa sổ trọng số được áp dụng với các đoạn trước khi chúng được ghép nối với đoạn đầu chồng lấp với nhau để tạo ra sự chuyển tiếp mịn giữa đoạn ghép nối được khôi phục ở đầu ra. Để xử lý thoại, WSOLA có ưu điểm là bảo toàn chu kỳ "pitch", đây là yêu cầu khắt khe để đảm bảo về mặt chất lượng âm thoại.

Cải tiến thuật toán WSOLA cho từng gói

Do tính tự nhiên của mất gói và hoạt động đa gói, trễ với khoảng thời gian 2 đến 3 gói được xem xét cho việc dãn gói. Vì mục tiêu của đọc ra thích nghi là để giảm bớt trễ, trễ xử lý thấp hơn là điều mong muốn. Do đó chúng ta phải biến đổi thuật toán WSOLA và cải tiến nó thực hiện với mỗi gói. Mặt khác, một gói đến có thể được xác lập tỷ lệ co dãn ngay lập tức và độc lập mà không làm tăng thêm trễ xử lý. Để xác lập tỷ lệ co dãn một gói, trước tiên chúng ta chọn một đoạn mẫu có chiều dài cố định ở đầu vào và sau đó tìm kiếm đoạn đồng dạng với đoạn mẫu. Đoạn đồng dạng bắt đầu được tìm kiếm trong vùng tìm kiếm như trong hình 6.16. Khi thực hiện trên từng gói, việc tìm kiếm cho một đoạn đồng dạng được chặt chẽ hơn, vì sự tập hợp lại các đoạn đồng dạng phải được thực hiện trong các khối của chu kỳ "pitch" và có ít hơn các chu kỳ "pitch" trong một gói ngắn. Với chiều dài một gói là 20 ms (ví dụ chuẩn G.729) và tùy thuộc vào giới tính của người nói, có thể có ít hơn hai chu kỳ "pitch", điều này làm cho nó khó trích ra đoạn đồng dạng. Để khắc phục vấn đề này, chúng ta thay đổi thuật toán để giảm chiều dài đoạn cho tính toán tương quan và vị trí đoạn mẫu đầu tiên bắt đầu của gói vào, như thể hiện trên hình 6.16(a). Để kéo dãn các gói ngắn, chúng ta di chuyển vùng tìm kiếm cho đoạn đồng dạng đầu tiên tới gói trước đó để có phạm vi rộng hơn tìm kiếm các dạng sóng đồng dạng. Trong hình 6.16(a), mặc dù gói đầu vào bắt đầu trong chu kỳ pitch thứ 2, đoạn đồng dạng được tìm thấy trong chu kỳ của pitch thứ nhất. Mặc dù gói trước đó có thể sẵn sàng đọc ra ở thời điểm xác lập thang thời gian, sóng đồng dạng có thể vẫn được trích ra từ nó để cấu trúc cho đầu ra mới mà không gây trễ gói trước đó. Một khi đoạn đồng dạng được tìm thấy,

trọng số được tăng thêm bởi mở rộng cửa sổ và tăng thêm trọng số đoạn mẫu bởi cửa sổ dốc đối xứng. Đoạn đồng dạng tiếp theo bởi phần còn lại của mẫu trong gói tiếp theo sau đó được di chuyển và được xếp chồng lên với đoạn mẫu để tạo ra dạng sóng cho đầu ra. Kết quả đầu ra dài hơn đầu vào do vị trí tương đối của đoạn đồng dạng tìm thấy và sự dịch chuyển của đoạn đồng dạng như trình bày trên hình 6.16(a). Ý nghĩa của khai triển phụ thuộc vào vị trí và cỡ của vùng tìm kiếm được định nghĩa.



Hình 6.16: (a) dãn và (b) nén từng gói sử dụng TSM

Trong hình 6.16 giữa các chu kỳ "pitch" trong đoạn sóng được phân chia bởi đường nét đứt thẳng đứng và được đánh dấu với các số thứ tự. Ví dụ trong hình 6.16, chúng ta có thể quan sát từ dạng sóng đưa ra mà một chu kỳ pitch bổ sung được tạo ra và bổ sung như kết quả của sự tập hợp lại và chồng lấp của các đoạn được trích ra từ đầu vào. Tuy nhiên, chu kỳ "pitch" mở rộng không chỉ là sự lặp lại của chu kỳ "pitch" nào từ đầu vào mà là nội suy từ một vài chu kỳ pitch.

Để tìm kiếm đoạn đồng dạng của đoạn mẫu trong vùng tìm kiếm, sử dụng hàm tương quan để tính toán giữa đoạn mẫu (*TemplateBlock*) và đoạn nguồn (*SourceBlock*) trong miền tìm kiếm. Mỗi đoạn nguồn trong miền tìm kiếm đồng dạng được dịch đi δ, trong đó $-L/2 \le \delta \le -L/2-1$ với L là chiều dài của miền tìm kiếm.

Đánh giá mức độ đồng dạng sử dụng hàm tương quan chéo như công thức (6.9) dưới đây.

$$\rho(\delta) = \sum_{k=0}^{N} \text{TemplateBlock}(k) \times \text{SourceBlock}(k+\delta) \quad (6.9)$$

Trong đó, N là chiều dài của một đoạn. Cửa sổ trọng số sử dụng cho tìm kiếm là cửa sổ cosin tăng (raised-cosine), $h(n) = 0{,}5 - 0{,}5\cos(2\pi n/N)$. Cỡ của cửa sổ được thiết lập là chiều dài của đoạn N.

Chất lượng thoại và thời gian tính toán của thuật toán phụ thuộc vào cỡ đoạn và chiều dài của vùng tìm kiếm cho đoạn nguồn. Các đoạn lớn sẽ chứa nhiều chu kỳ pitch nên việc tìm tương quan đồng dạng giữa đoạn mẫu và đoạn đồng dạng tốt

hơn. Tuy nhiên, nếu cỡ đoạn quá lớn sẽ sinh tiếng vọng và các âm thanh lạ ở đầu ra. Vùng tìm kiếm lớn cũng sẽ cho kết quả tương quan tốt hơn nhưng cũng phải tính toán nhiều hơn.

Trong hình 6.16(a) ba chu kỳ pitch đầu tiên là sự chồng lấp có trọng số của các chu kỳ pitch 1/2, 2/3, và 3/4 tương ứng. Điều này giải thích tại sao chất lượng âm thanh sử dụng kỹ thuật co dãn thời gian là tốt hơn lặp lại "pitch" (như trình bày trong chương trước). Tương tự như thế cho việc nén gói, trong đó thông tin được mang bởi chu kỳ "pitch" được chặt nhỏ để bảo tồn và phân bổ trong số chu kỳ còn lại. Tuy nhiên, thực hiện với từng gói mô tả trên có ưu điểm giống như lặp "pitch" là không làm tăng thêm trễ.

Nếu thoại đầu ra không tiệm cận được chiều dài mong đợi sau một số tính toán thì sự lặp lại bổ sung được thực hiện. Trong một sự lặp lại tiếp theo, một đoạn mẫu mới của cùng chiều dài được định nghĩa ngay lập tức theo mẫu trong đoạn lặp lại cuối cùng. Tất cả các đoạn mẫu được định nghĩa và mẫu còn lại sau mẫu cuối cùng trong đầu vào để bao trùm toàn bộ đầu ra với chiều dài cần đạt được. Con số tổng của các đoạn mẫu được xác định và số của đoạn lặp được sử dụng ở đây là:

$$\left\lfloor \frac{\hat{L}^{i}}{W} \right\rfloor - 1 \qquad (6.10)$$

Trong đó $[x]$ biểu diễn cho số tự nhiên lớn nhất mà nhỏ hơn hoặc bằng x, $\hat{L}^{i}$ là chiều dài cần đạt được của đầu ra và W là chiều dài của đoạn (hoặc là đoạn mẫu hoặc là đoạn đồng dạng).

Nén gói được thực hiện theo cách tương tự như mô tả trong hình 6.14(b). Chỉ một điểm khác đó là vùng tìm kiếm cho đoạn đồng dạng sẽ không được định nghĩa trong gói trước đó để tạo ra ở đầu ra ngắn hơn chiều dài. Nén gói thành công yêu cầu một gói chứa đựng nhiều hơn một "pitch", nó giới hạn chiều dài tối thiểu của gói có thể được nén. Tuy nhiên, chiều dài một gói phổ biến cỡ 20 ms thường hiệu quả vì giá trị "pitch" dưới 100 Hz là không thường xuyên trong tín hiệu thoại. Cũng có thể trong một vài trường hợp, nén gói không thể được thực hiện, nó không thích hợp với trễ nén với các gói tiếp sau đó.

So sánh dạng sóng đầu vào và đầu ra trong hình 6.14, nó hiển nhiên là bảo đảm hoạt động tần số "pitch" của thoại đầu vào. Chỉ có chiều dài các gói thay đổi và do đó tốc độ thoại được biến đổi.

Một điểm lợi của kỹ thuật này với một gói ngắn đó là đầu vào được chia thành ít các đoạn mẫu hơn để chỉ một hoặc hai đoạn lặp đặc trưng sẽ cho một đầu ra có chiều dài như mong đợi. Đặc điểm quan trọng khác của thuật toán thể hiện rõ trong hình 6.14 đó là bắt đầu và kết thúc của mỗi gói không bị biến đổi. Như kết quả, khi các gói được sửa đổi tái tạo được nối vào mà không gây chồng chéo lên nhau bảo đảm sự chuyển đổi dạng sóng tự nhiên và trơn mượt. Do đó, các gói được sửa đổi độc lập và gửi đi đến hàng đợi đầu ra. Kiểu thực hiện này là rất phù hợp cho bộ lập chương trình đọc ra thích nghi cao.

Thực hiện tính toán thời gian đọc ra

Vì co dãn các tỷ lệ các gói đã được thực hiện theo bội số tự nhiên của các chu kỳ "pitch", nó không thể thực hiện chiều dài tùy ý gói và thời gian đọc ra như được mong đợi cho đọc ra thích nghi. Mặt khác, chiều dài gói tính toán được thực tế L_i sau một gói WSOLA có thể chỉ xấp xỉ chiều dài được yêu cầu $\hat{L}^{i}$. Vì lý do này, chúng ta định nghĩa "ngưỡng dãn" và "ngưỡng nén". Nếu chỉ thời gian đọc mong muốn lớn hơn thời gian đọc được lập trình do vượt quá ngưỡng nén, chúng ta sẽ nén một gói để tăng tốc độ đọc ra. Ngưỡng nén thường lớn hơn một chu kỳ pitch điển hình. Sử dụng cùng phương pháp để dãn một gói, trừ phi hai ngưỡng nén là không đối xứng. Để ngăn chặn mất gói do trễ không cần thiết, chúng ta áp dụng nén từ từ đủ để tránh rớt dưới thời gian đọc ra. Ngưỡng dãn nhỏ hơn được định nghĩa có thể nhỏ hơn một chu kỳ “pitch”. Trong trường hợp này chúng ta dãn gói và giảm tốc độ đọc ra để làm thích nghi với hiện tượng tăng trễ mạng. Kết quả không đối xứng này có thể thấy được trong hình 6.13(c), trong đó trễ mạng tăng dễ bám hơn là khi trễ mạng giảm. Trình bày hiện tượng trễ cũng như kết quả trong biến động đọc ra được làm trơn mượt.

Để tránh tỷ lệ co dãn lớn, chúng ta cũng định nghĩa chiều dài gói xác định tối đa và tối thiểu, ký hiệu tương ứng bởi *Lmax* và *Lmin*. Tuy nhiên, trong suốt chu kỳ khoảng lặng, giá trị điều chỉnh có thể làm chương trình đọc ra không bị giới hạn bởi *Lmax* hoặc *Lmin*, cho chúng ta thoải mái hơn để thay đổi chương trình đọc ra. Thủ tục chung của việc điều chỉnh chương trình đọc ra được mô tả bởi thuật toán trong bảng 6.4.

Bảng 6.4: Thuật toán điều chỉnh thời gian đọc ra thích nghi với trễ mạng

Nhận gói i;

Ước lượng và tính thời gian đọc ra của gói i + 1, $\hat{t}_p^{i+1}$;

Tính toán chiều dài mong đợi của gói i, $\hat{L}^i = \hat{t}_p^{i+1} - t_p^i$;

if $\hat{L}^i - L_0$ > ngưỡng dãn

Co dãn gói i với chiều dài tính toán min($\hat{L}^i$, Lmax);

elseif $\hat{L}^i - L_0$ < -ngưỡng nén

Co dãn gói i với chiều dài tính toán max($\hat{L}^i$, Lmin);

else

Giữ nguyên gói i không cần thay đổi chiều dài;

endif

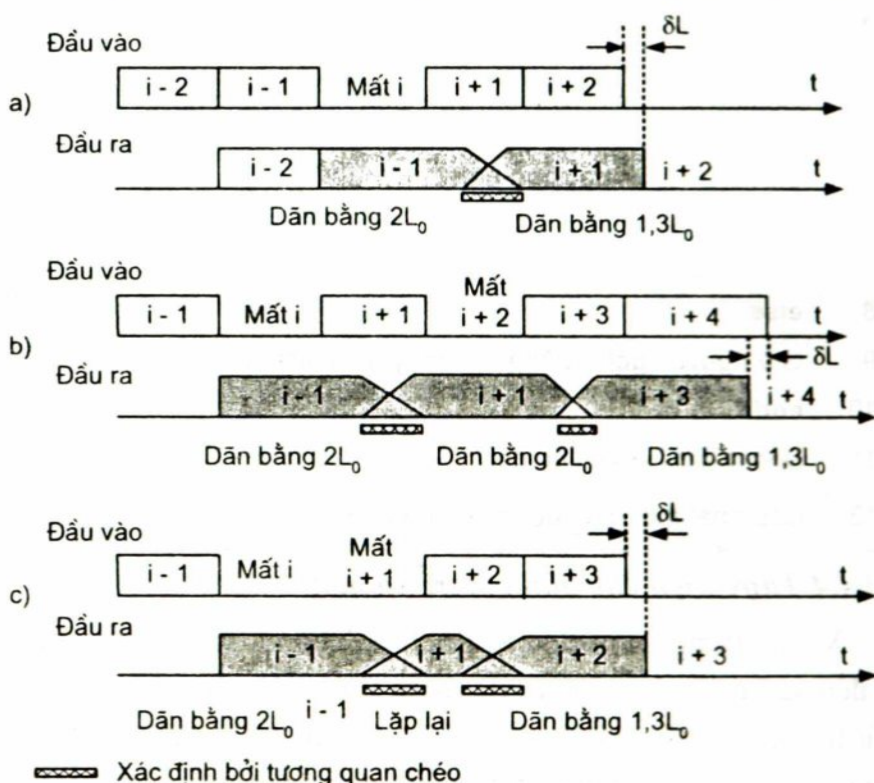
Đọc ra gói i với chiều dài L^i;

Cập nhật thời gian đọc ra của gói i + 1, $t_p^{i+1} = t_p^i + L^i$;

6.3.4.4 Thực hiện che dấu gói trễ và mất

Với chương trình đọc ra thích nghi, một số nào đó của gói sẽ đến sau thời gian đọc ra đã được tính toán cho chúng hoặc bị mất trên mạng. Để khôi phục thông tin mất tốt nhất có thể, khá nhiều kỹ thuật khôi phục mất gói đã nghiên cứu trước đây. Nó là sự ghép lai giữa kỹ thuật co dãn thang thời gian và thay thế dạng sóng, được sử dụng cho cả mất gói do trễ và mất do lỗi liên kết mạng bằng cách lợi dụng thông tin dư thừa trong tín hiệu audio.

Tuy nhiên, trễ thuật toán bằng khoảng thời gian đọc của 2 đến 3 gói. Phương pháp của chúng ta có ưu điểm là xác lập thang thời gian cho từng gói và sử dụng thông tin hai chiều do làm việc cùng với chương trình đọc ra thích nghi. Phương pháp che dấu này làm giảm trễ với thời gian một gói và kết quả là chất lượng thoại tốt hơn. Kỹ thuật lặp gói không đưa ra một trễ thuật toán nào ngoài thời gian chu kỳ ghép rất ngắn, tuy nó không đảm bảo chất lượng âm thanh tốt như kỹ thuật co dãn thang thời gian.



Hình 6.17: Che dấu mất gói, (a) mất từng gói; (b) mất xen kẽ; (c) mất liên tiếp

Phương pháp che dấu được đề xuất như minh họa trong hình 6.17. Để đơn giản cho các phân tích tiếp theo chúng ta bỏ qua thời gian xử lý của thời gian xác lập gói và giả sử rằng một gói có thể được co dãn ngay lập tức và sau đó được đọc ra. Giả sử rằng ở thời điểm gói i bị mất, gói trước đó i - 1 bị kéo dãn với chiều dài xác định là $2L_0$ và được đọc ra. Bởi vì phải đợi để nhận trạng thái của gói i, gói i - 1 phải nằm trong bộ đệm và bị trễ bằng thời gian của một gói. Gói i coi như bị mất nếu nó không nhận được tại thời điểm gói i - 1 được đọc ra và kỹ thuật che dấu mất gói bắt đầu được thực hiện.

Hơn nữa, phép toán thực hiện phụ thuộc vào gói bị mất. Nếu chỉ có gói i bị mất (tức là gói i + 1 và i + 2 nhận được trong thời gian hạn định của chúng), gói i + 1 được kéo dãn với một chiều dài xác định bằng $1,3L_0$. Trước ghép nối hai gói đã được dãn là i - 1 và i + 1, thì cần phải tìm kiếm một đoạn sóng đồng dạng trong đoạn giới hạn giữa cuối của gói i - 1 và đầu gói i + 1 (đoạn này khoảng bằng chiều dài một nửa chu kỳ pitch) trong vị trí của đoạn được ghép. Tìm kiếm đoạn sóng đồng dạng ở vị trí ghép chồng được thực hiện bằng hàm tương quan chéo. Như cho thấy trên hình 6.15(a), các đoạn ngắn được tìm thấy từ một trong hai phía được tăng thêm trọng số bằng việc giảm hoặc tăng cửa sổ tìm kiếm trước khi ghép. Bằng cách này đã tránh được các điểm gián đoạn (chuyển tiếp từ cuối gói i - 1 sang đầu gói i + 1) và làm chất lượng âm thanh tốt hơn so với phương pháp ghép cố định thông thường. Toàn bộ khoảng dãn của gói i - 1 và i + 1 sẽ bao trùm và gần như giống với dạng sóng gói bị mất.

Một điều nhận thấy rằng, không bảo đảm chính xác hoàn toàn về chiều dài của các gói ghép nối như khi dãn do cỡ đầu ra không được quyết định bởi WSOLA và thủ tục ghép đã mô tả. Chiều dài thu được của các gói đã co dãn bằng $L^{i-1} + L^{i+1} - L_{merge}$, và thời gian đọc ra của gói i + 2 là:

$$t_p^{i+2} = t_p^{i-1} + L^{i-1} + L^{i+1} - L_{merge} \quad (6.11)$$

Chú ý rằng che dấu thành công sẽ có $t_p^{i+2} > t_r^{i+2}$. Nhìn chung, t_p^{i+2} sẽ không tương ứng với thời gian đọc ra mong đợi. Do đó, thời gian đọc ra thực tế gần như nhanh hơn hoặc chậm hơn thời gian đọc ra được tính toán một khoảng sai lệch là δL, như minh họa trên hình 6.17. Tuy nhiên, điều này có thể hiệu chỉnh bởi thuật toán đọc ra thích nghi, giải quyết bằng cách thay đổi tỷ lệ co dãn các gói tiếp theo. Do đó, sử dụng kỹ thuật co dãn thang thời gian cho đọc ra thích nghi cũng như che dấu mất gói cho phép bổ sung tính mềm dẻo và dễ tích hợp trong hệ thống.

Ngoài mất gói đơn, phương pháp che dấu mất gói của chúng ta có thể điều khiển các mẫu mất được ghép xen hoặc mất cụm như cho thấy trên hình 6.17 (b) và (c). So sánh các kỹ thuật che dấu có sự cải thiện đáng kể có thể đạt được cho các trường hợp này. Ưu điểm chính của phương pháp đề xuất đó là nó có thể phát hiện các mẫu và điều chỉnh tự động tỷ lệ co dãn. Trong trường hợp khác, khi gói i bị mất, co dãn gói i - 1 tiếp theo cũng như thủ tục được mô tả trong hình 6.17 (a). Vì mẫu sẽ mất là không biết ở thời điểm i. Trong hình 6.17 (b) gói i + 2 được quyết định là bị

mất, gói i + 1 bị dãn với chiều dài xác định là $2L_0$ thay vì $1,3L_0$, để bao trùm khoảng thời gian trống cho gói bị mất thứ hai và gói i + 3 sẽ dãn với chiều dài là $1,3L_0$. Trong hình 6.17 (c), các gói i và i + 1 bị mất, gói i - 1 được kéo dãn ra với chiều dài $2L_0$ và dạng sóng của gói đã dãn i - 1 được lặp lại để che dấu mất cụm. Trong cả hai trường hợp, tìm kiếm các dạng sóng đồng dạng được thực hiện để ghép và điều chỉnh thời gian đọc ra thích nghi được sử dụng cho các gói tiếp theo nếu cần thiết.

Cuối cùng, che dấu hai hoặc nhiều hơn các gói bị mất liên tiếp minh hoạ trên hình 6.17(c). Khi nhiều hơn hai gói bị mất liên tiếp xảy ra, lặp lại dạng sóng được sử dụng tối đa là ba lần, trước khi cơ chế dừng để tạo ra dạng sóng ở đầu ra và xác lập lại chương trình đọc ra. Do tính tái tạo tự nhiên của lặp dạng sóng, mất cụm giảm chất lượng thoại, ngay cả sau khi được che dấu. Mặc dầu các cơ chế tái tạo hiện thời sử dụng ghép xen hoặc FEC (như đã trình bày trong mục 6.2 của chương) để bảo vệ dữ liệu khỏi mất cụm, các phương pháp này được thực hiện với giá phải trả là trễ cao. Như cho thấy, một ưu điểm đặc biệt của chương trình đọc ra thích nghi là khả năng giảm mất các gói liên tiếp như mất cụm.

Thuật toán được đề nghị cho phương pháp che dấu mất gói trình bày trong bảng 6.5. Phương pháp này đưa ra trễ thời gian bằng một gói nhưng chất lượng thoại tốt và xử lý tỷ lệ mất gói cao của nhiều loại mẫu khác nhau. Chú ý rằng khi được sử dụng cùng với chương trình đọc ra thích nghi, thời gian đọc ra trong bảng 6.4 sẽ được bù bằng một gói. Do các phương pháp che dấu

khác cũng có thể được sử dụng, thuật toán cho chương trình đọc ra thích nghi trong thuật toán bảng 6.4 được xem xét mà không nhất thiết sử dụng cơ chế che dấu mất gói cụ thể nào. Hơn nữa, nếu một bộ mã hóa thoại được sử dụng để truyền đó là thuật toán che dấu nội tại hoạt động với trễ thấp, nó có thể thuận lợi để sử dụng thuật toán được đề xuất ở đây. Nhìn chung, ta có thể tin rằng cơ chế đề xuất có thể tích hợp tốt với bất kỳ kiểu bộ mã hóa thoại nào.

Bảng 6.5: Thuật toán sử dụng cho che dấu mất gói

```
if gói i nhận được
    if gói i - 1 nhận được
      if gói i - 2 nhận được
        Đọc ra gói i - 1 với chiều dài L_{i-1};
      else
        Dãn gói i - 1 với chiều dài xác định là 1.3L_0;
        Đọc ra gói i - 1;
      endif
    endif
else
    if gói i - 1 nhận được
        Dãn gói i - 1 với chiều dài xác định là 2L_0;
        Đọc ra gói i - 1;
    else
        Lặp lại và đọc ra gói i - 2 sử dụng lặp dạng sóng;
endif
endif
```

Như trình bày ở trên, thuật toán đọc ra thích nghi trong bảng 6.4 sẽ xác định chiều dài và co dãn gói đọc ra, nhưng khi mất gói

thuật toán che dấu mất gói trong bảng 6.5 được sử dụng. Nó sẽ quyết định "lấp đầy" vào gói mất bằng cách dãn các gói lân cận với tỷ lệ dãn nào đó hay lặp lại một đoạn sóng đồng dạng nội suy từ gói trước đó.

Tóm lại: Trong phần chương này, trình bày kỹ thuật co dãn theo thời gian sử dụng thuật toán đọc ra thích nghi và thuật toán WSOLA để thực hiện việc khôi phục, hay cụ thể hơn là che dấu các gói bị mất và thuật toán đọc ra thích nghi hạn chế đến mức tối đa mất gói khi truyền thoại qua mạng IP. Vì phương pháp đề xuất chỉ thực hiện ở đầu thu. Không yêu cầu chức năng hỗ trợ gì từ phía mạng hoặc bộ phát nên có khả năng áp dụng cao. Cơ chế ước lượng trễ mạng dựa trên thuật toán ước lượng sử dụng lọc thích nghi NLMS thích nghi tốt với biến động nhanh của trễ mạng và đặc biệt trễ đỉnh. Thuật toán đọc ra thích nghi sẽ trễ mạng ước lượng được để tính toán thời gian đọc ra cho từng gói. Mỗi gói sẽ có thời gian đọc ra phụ thuộc vào trễ mạng của các w gói trước đó với w là độ rộng cửa sổ tức là w gói trước gói thứ i đang xét. w lớn có độ chính xác cao hơn nhưng phải tính toán nhiều và sẽ thích nghi chậm hơn trong trường hợp trễ đỉnh.

Do thời gian đọc từng gói ra không theo chu kỳ như khi phát nên phải sử dụng kỹ thuật TSM để co dãn gói đọc ra (nghĩa là chiều dài gói lớn hoặc nhỏ hơn chiều dài đoạn thoại đóng gói), thuật toán WSOLA được cải tiến áp dụng để thực hiện co dãn cho từng gói mà vẫn đảm bảo chất lượng âm thanh, không tăng thêm trễ và không gây gián đoạn ở điểm tiếp giáp giữa các gói nhờ sử dụng tìm kiếm bằng tương quan chéo. WSOLA cũng được áp dụng cho trường hợp cần kéo dãn các gói để che dấu

cho vị trí gói bị mất và kết hợp sử dụng kỹ thuật lặp sóng đồng dạng trong trường hợp mất cụm xảy ra như trong hình 6.17 bằng thuật toán che dấu mất gói.

Tỷ lệ co dãn được xác định với chiều dài gói $0{,}5L_0$, $0{,}7L_0$, 1,3 L_0, $2L_0$ hoặc giữ nguyên chiều dài gói (L_0). Các kết quả nghiên cứu đã cho thấy tỷ lệ co dãn trên vẫn đảm bảo chất lượng tốt.

6.4. GIẢI PHÁP NÂNG CAO CHẤT LƯỢNG MẠNG

6.4.1 Bổ sung QoS vào mạng

Nếu một mạng thiếu QoS thì người sử dụng phải tự thêm vào các ứng dụng của mình để có được QoS cần thiết. Việc thêm bảo mật vào ở lớp ứng dụng đã được đề cập đến ở phần trước. Khi thực hiện VoIP trên mạng Internet công cộng, các bộ đệm *biến động trễ* thường được sử dụng tại phía bộ thu.

Trong tất cả các thông số QoS, thông số mà người sử dụng khó tự giải quyết nhất là trễ. Mà thoại lại là ứng dụng rất nhạy cảm với trễ mạng.

Tuy nhiên, nếu người sử dụng không sẵn sàng chi tiền để thêm QoS cho mỗi ứng dụng, thì nhà cung cấp dịch vụ nên làm gì để cải thiện QoS của mạng? Người sử dụng không thể hạ thấp trễ mạng, nhưng nhà cung cấp dịch vụ có nên quan tâm tới việc đầu tư làm giảm trễ trên toàn bộ mạng khi mà chủ yếu các ứng dụng yêu cầu trễ thấp như là thoại tạo ra lợi nhuận? Tại sao không đầu tư xây dựng một mạng khác? Đây là lí do vì sao các mạng thoại và dữ liệu cùng tồn tại riêng rẽ lâu đến thế.

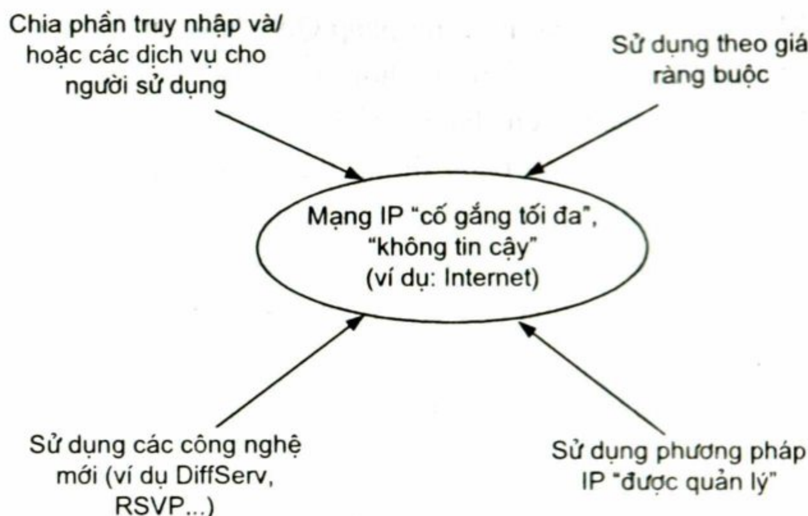
Một lý do khác cho phương pháp QoS "chỉ cần băng thông". Thực tế đơn giản là sẽ hiệu quả hơn nếu khắc phục nhược điểm của QoS bằng cách thêm băng thông hơn là bằng bất cứ cách nào khác. Thêm vào đủ băng thông thì ít nhất trễ và *biến động trễ* sẽ được cải thiện. Nếu băng thông được thêm vào đúng đắn, thì thậm chí cả vấn đề mất gói, tính sẵn sàng và bảo mật cũng sẽ được cải thiện. Hiệu quả của việc thêm băng thông là có thể biết trước được.

Xây dựng một mạng có QoS hay bổ sung QoS vào mạng có sẵn?

Như vậy có hai lựa chọn để nhận được QoS cho người sử dụng và cho ứng dụng, đó là xây dựng một mạng có QoS hoặc là thêm QoS vào trong mạng có sẵn. Điều này thực chất không có gì khác hơn là cân nhắc giữa việc có sẵn QoS ở trong mạng và việc thêm QoS ở cấp ứng dụng ở bên ngoài mạng.

Hai phương pháp này đã được thảo luận ở phần trước. Phương pháp thứ nhất, trong đó người sử dụng nói với mạng họ muốn gì, là phương pháp được thực hiện bởi hầu hết các nhà cung cấp dịch vụ PSTN. Phương pháp thứ hai, trong đó người sử dụng thêm QoS của chính họ và chỉ tin tưởng mạng không gì hơn là ở các kết nối, đây chính là triết lý của Internet.

Đối với các nhà cung cấp dịch vụ ngoài cách đơn giản là thêm băng thông thì chỉ có ba cách để giải quyết nghẽn và thiếu tài nguyên trên mạng. Và chỉ một trong các cách này cũng đòi hỏi phải tiêu tốn một số tiền lớn để thêm một thông số QoS cụ thể vào một mạng như mạng Internet. Các phương pháp này được đưa ra trong hình 6.18.



Hình 6.18: Bổ sung QoS vào mạng

Một nhà cung cấp dịch vụ có thể:

- *Chia phần truy nhập và/hoặc các dịch vụ cho người sử dụng*. Phương pháp này đòi hỏi nhà cung cấp dịch vụ phải kiểm soát được mạng và biết được ai đang sử dụng phần tài nguyên nào của mạng. Hiện tại không có cách nào dễ dàng để thực hiện điều này trên Internet, trừ cách cố gắng xác định các mô-đem quay số rỗi và đưa chúng ra ngoài và đương nhiên phương pháp này không được ưa chuộng trong cộng đồng người sử dụng.

- *Sử dụng theo giá ràng buộc và/hoặc quản lý IP*. Phương pháp này buộc người sử dụng phải trả nhiều tiền hơn để sử dụng nhiều tài nguyên mạng hơn. Các tuyến tĩnh có thể được thiết lập riêng cho kết nối các điểm sử dụng VoIP để tối thiểu trễ, nhưng sẽ phải trả giá cao hơn phí kết nối ISP. Sử dụng theo chi phí ràng

buộc bao gồm các bộ định tuyến IP được quản lý hoàn toàn và các liên kết dành riêng cho một hoặc một số giới hạn các khách hàng lớn.

- *Sử dụng công nghệ mới.* Người sử dụng không thích chia phần truy nhập, và khách hàng không nhất thiết phải tiêu một lượng tiền lớn để quản lý mạng IP. Cách đạt được QoS cuối cùng là sử dụng một hoặc nhiều công nghệ mới để làm cho mạng IP trở nên tốt như mạng PSTN khi cung cấp thoại. Câu hỏi đặt ra là công nghệ mới nào? Có nhiều công nghệ và chúng sẽ được xem xét.

Người sử dụng có thể làm gì với một mạng IP để cung cấp QoS của riêng họ nếu QoS đầy đủ cho ứng dụng không sẵn sàng từ một ISP? Bảng 6.6 liệt kê một số điều có thể làm để thêm các thông số QoS từ một ISP.

Bảng 6.6. Thêm QoS ứng dụng vào một mạng IP

Để cải thiện thông số QoS ...	**Người sử dụng có thể ...**
Băng thông	Dồn kênh theo hướng ngược để tăng dung lượng cho nhiều cuộc gọi
Trễ	Không làm gì nhiều trừ cố gắng tối thiểu các bước nhảy giữa các bộ định tuyến (nhưng làm thế nào?)
Biến động trễ	Thêm các bộ đệm biến động trễ
Mất thông tin	Thêm phần sửa lỗi trước vào gói thoại (ít dùng)
Tính sẵn sàng	Sử dụng nhiều liên kết đến ISP, thậm chí sử dụng nhiều ISP
Bảo mật	Thêm các phương pháp nhận thực và mã hóa của chính họ (hay dùng)

Hầu hết các hành động của người sử dụng đều không cần giải thích gì thêm, nhưng nên nhớ rằng những hành động được áp dụng trong chương này chủ yếu liên quan đến VoIP. Hiếm khi thoại số 4 kbit/s hoặc 8 kbit/s gây ra vấn đề về băng thông cho người sử dụng trên Internet, trừ trường hợp mạng bị nghẽn nặng. Dồn kênh theo hướng ngược có thể cần để cung cấp nhiều luồng lưu lượng VoIP thay vì kết nối VoIP riêng lẻ. Các bộ đệm *biến động trễ* được sử dụng rất rộng rãi với bất kỳ hình thức VoIP nào không cần biết có bao gồm Internet hay không, bởi vì chỉ có PSTN trên cơ sở kênh mới có giới hạn đủ chặt chẽ về *biến động trễ* để thoả mãn thoại. Ảnh hưởng của lỗi có thể được tối thiểu một phần bằng cách sử dụng mã FEC (Forward Error-Correcting Code), nhưng điều này hiếm khi được thực hiện. FEC yêu cầu tính tương thích nên do đó giới hạn sự lựa chọn của người sử dụng, và FEC cũng không giúp được gì khi toàn bộ gói bị mất trên Internet. Tính sẵn sàng được tăng cường bằng cách sử dụng nhiều liên kết tới một ISP hoặc thậm chí sử dụng nhiều ISP. Cuối cùng, bảo mật theo truyền thống là một vấn đề cần sự quan tâm về QoS của người sử dụng và VoIP hay điện thoại Internet có thể sử dụng một trong nhiều cách để đảm bảo tính riêng tư và nhận thực, mặc dù khả năng hoạt động cùng với nhau vẫn còn là một vấn đề cần giải quyết.

6.4.2. Các mô hình cung cấp chất lượng dịch vụ

Mục đích của việc cung cấp chất lượng dịch vụ trong mạng IP là để đảm bảo việc truyền gói và phân biệt các dịch vụ trong mạng. Có nhiều mức khác nhau để đảm bảo và phân biệt các

dịch vụ. Mỗi mức đó đại diện cho một mô hình cung cấp chất lượng dịch vụ.

Khi đề cập đến VoIP, có nhiều điều cần xác định với QoS hơn là sáu thông số QoS như được đặt ra ở chương trước. Có những vấn đề nổi lên chỉ đơn giản bởi vì Internet và tất cả các mạng IP về bản chất đều là tập hợp của các bộ định tuyến phi kết nối, chúng định tuyến mỗi gói tin độc lập với nhau chỉ trên cơ sở các thông tin mạng cục bộ. Theo đó, một bộ định tuyến có thể gửi một gói tin tới một bộ định tuyến liền kề gần đích hơn. Thông tin toàn cục mà bộ định tuyến chia sẻ chỉ là thông tin kết nối cơ bản, trừ khi các bảng đặc biệt được xây dựng và tham khảo, điều này sẽ làm chậm quá trình định tuyến.

Ngoài trễ, biến động trễ và mất gói, thì trong một mạng sử dụng VoIP chỉ còn các vấn đề về bộ định tuyến là cần quan tâm. Trong thực tế, mất gói trên một mạng IP do nhiều nguyên nhân, không chỉ từ ảnh hưởng của lỗi trên liên kết hay do nghẽn trong bộ định tuyến. Tiêu đề gói IP chứa trường *thời gian tồn tại* (TTL), trường này giới hạn số bước nhảy (số bộ định tuyến) mà một gói tin có thể đi qua trước khi đến được đích. Mỗi bộ định tuyến sẽ giảm biến đếm TTL trước khi xử lý gói. Nếu trường TTL bằng 0 sau khi giảm thì gói sẽ bị loại bỏ và bên gửi sẽ được thông báo. Nhưng điều này không giúp được gì nhiều cho các ứng dụng thời gian thực như VoIP, vì các ứng dụng này không thể gửi lại gói.

Một mô hình dịch vụ còn được gọi là một mức dịch vụ mô tả khả năng thiết lập từ đầu cuối đến đầu cuối của QoS, đầu cuối

Hầu hết các hành động của người sử dụng đều không cần giải thích gì thêm, nhưng nên nhớ rằng những hành động được áp dụng trong chương này chủ yếu liên quan đến VoIP. Hiếm khi thoại số 4 kbit/s hoặc 8 kbit/s gây ra vấn đề về băng thông cho người sử dụng trên Internet, trừ trường hợp mạng bị nghẽn nặng. Dồn kênh theo hướng ngược có thể cần để cung cấp nhiều luồng lưu lượng VoIP thay vì kết nối VoIP riêng lẻ. Các bộ đệm *biến động trễ* được sử dụng rất rộng rãi với bất kỳ hình thức VoIP nào không cần biết có bao gồm Internet hay không, bởi vì chỉ có PSTN trên cơ sở kênh mới có giới hạn đủ chặt chẽ về *biến động trễ* để thoả mãn thoại. Ảnh hưởng của lỗi có thể được tối thiểu một phần bằng cách sử dụng mã FEC (Forward Error-Correcting Code), nhưng điều này hiếm khi được thực hiện. FEC yêu cầu tính tương thích nên do đó giới hạn sự lựa chọn của người sử dụng, và FEC cũng không giúp được gì khi toàn bộ gói bị mất trên Internet. Tính sẵn sàng được tăng cường bằng cách sử dụng nhiều liên kết tới một ISP hoặc thậm chí sử dụng nhiều ISP. Cuối cùng, bảo mật theo truyền thống là một vấn đề cần sự quan tâm về QoS của người sử dụng và VoIP hay điện thoại Internet có thể sử dụng một trong nhiều cách để đảm bảo tính riêng tư và nhận thực, mặc dù khả năng hoạt động cùng với nhau vẫn còn là một vấn đề cần giải quyết.

6.4.2. Các mô hình cung cấp chất lượng dịch vụ

Mục đích của việc cung cấp chất lượng dịch vụ trong mạng IP là để đảm bảo việc truyền gói và phân biệt các dịch vụ trong mạng. Có nhiều mức khác nhau để đảm bảo và phân biệt các

dịch vụ. Mỗi mức đó đại diện cho một mô hình cung cấp chất lượng dịch vụ.

Khi đề cập đến VoIP, có nhiều điều cần xác định với QoS hơn là sáu thông số QoS như được đặt ra ở chương trước. Có những vấn đề nổi lên chỉ đơn giản bởi vì Internet và tất cả các mạng IP về bản chất đều là tập hợp của các bộ định tuyến phi kết nối, chúng định tuyến mỗi gói tin độc lập với nhau chỉ trên cơ sở các thông tin mạng cục bộ. Theo đó, một bộ định tuyến có thể gửi một gói tin tới một bộ định tuyến liền kề gần đích hơn. Thông tin toàn cục mà bộ định tuyến chia sẻ chỉ là thông tin kết nối cơ bản, trừ khi các bảng đặc biệt được xây dựng và tham khảo, điều này sẽ làm chậm quá trình định tuyến.

Ngoài trễ, biến động trễ và mất gói, thì trong một mạng sử dụng VoIP chỉ còn các vấn đề về bộ định tuyến là cần quan tâm. Trong thực tế, mất gói trên một mạng IP do nhiều nguyên nhân, không chỉ từ ảnh hưởng của lỗi trên liên kết hay do nghẽn trong bộ định tuyến. Tiêu đề gói IP chứa trường *thời gian tồn tại* (TTL), trường này giới hạn số bước nhảy (số bộ định tuyến) mà một gói tin có thể đi qua trước khi đến được đích. Mỗi bộ định tuyến sẽ giảm biến đếm TTL trước khi xử lý gói. Nếu trường TTL bằng 0 sau khi giảm thì gói sẽ bị loại bỏ và bên gửi sẽ được thông báo. Nhưng điều này không giúp được gì nhiều cho các ứng dụng thời gian thực như VoIP, vì các ứng dụng này không thể gửi lại gói.

Một mô hình dịch vụ còn được gọi là một mức dịch vụ mô tả khả năng thiết lập từ đầu cuối đến đầu cuối của QoS, đầu cuối

đến đầu cuối QoS là khả năng của mạng có thể phục vụ các yêu cầu đặc biệt tới mạng khác. Có 3 kiểu mô hình dịch vụ:

- Cố gắng tối đa (best effort)

- Dịch vụ tích hợp (*IntServ* – Integrated Service)

- Dịch vụ khác biệt (*DiffServ* - Differentiated Service).

6.4.2.1 Cung cấp thêm tài nguyên mạng - Dịch vụ cố gắng tối đa (best-effort)

Cố gắng tối đa là một mô hình dịch vụ đơn cho phép ứng dụng gửi dữ liệu bất cứ khi nào với bất cứ khối lượng nào nó có thể thực hiện và không đòi hỏi sự cho phép hoặc thông tin cơ sở mạng, nghĩa là mạng phân phối dữ liệu nếu nó có thể mà không cần sự đảm bảo về độ tin cậy, độ trễ hoặc khả năng thông mạng.

Đặc tả dịch vụ cố gắng tối đa là xếp hàng đợi: vào trước, ra trước (first-in, first-out - FIFO). Dịch vụ cố gắng tối đa rất phù hợp cho những ứng dụng của mạng dải rộng như truyền tệp hoặc email.

Giải pháp có thể nghĩ tới để giải quyết vấn đề tắc nghẽn trên đường truyền mạng là cung cấp thêm tài nguyên mạng nhằm mục đích tăng giải thông của mạng trong trường hợp xảy ra tắc nghẽn. Tuy nhiên rõ ràng là giải pháp này không có tính kinh tế, ít nhất là đối với cấu trúc mạng như hiện nay, bởi vì chúng ta hiếm khi có thể dự đoán trước được thời điểm sẽ xảy ra tắc nghẽn mạng và cũng không thể tìm đủ tài nguyên mạng để thỏa mãn tất cả các nhu cầu trên mạng. Hơn nữa, dịch vụ theo kiểu cố gắng tối đa không thể đảm bảo cung cấp các dịch vụ tin cậy. Thậm chí ngay cả đối với những mạng không chịu tải lớn, thì độ

trễ trên mạng cũng đã đủ để ảnh hưởng tới chất lượng của các ứng dụng thời gian thực. Bởi vậy, để có thể cung cấp được những bảo đảm về chất lượng. Các dịch vụ IP phải hỗ trợ khả năng cung cấp nhiều dịch vụ khác nhau và nhiều mức độ dịch vụ cho từng lớp người dùng và ứng dụng khác nhau.

Mô hình cố gắng tối đa như chúng ta đã biết và cũng không phải trình bày nhiều, nó chính là mô hình cung cấp dịch vụ của mạng Internet hiện nay.

6.4.2.2 Mô hình lưu giữ tài nguyên – Cấu trúc dịch vụ tích hợp (IS - Integrated Service)

Cấu trúc dịch vụ tích hợp được định nghĩa bởi IETF nhằm mục đích đưa Internet trở thành một cơ sở hạ tầng cho các dịch vụ tích hợp tính năng cao hỗ trợ truyền thông tiếng nói, hình ảnh, dữ liệu thời gian thực cũng như các dữ liệu truyền thống. Các tài nguyên mạng sẽ được phân chia tùy theo các yêu cầu QoS cụ thể, và được tập trung vào các chiến lược quản lý dải thông. Các cơ chế để giải quyết vấn đề này được cung cấp bởi giao thức lưu giữ tài nguyên RSVP như là một phần của cấu trúc IntServ.

Hơn nữa, dịch vụ tích hợp là mô hình đa dịch vụ, nó có thể cung cấp nhiều yêu cầu QoS. Mô hình này đòi hỏi một loại hình đặc biệt của dịch vụ mạng trước khi nó gửi dữ liệu. Kiến trúc của mô hình được mô tả trong RFC 1633.

Những yêu cầu được tạo ra bằng việc báo hiệu chính xác. Ứng dụng đó thông tin tới mạng của nó về hiện trạng lưu lượng và những yêu cầu liên quan đến một loại hình dịch vụ đó, có thể

là các thông số về độ rộng băng, độ trễ. Ứng dụng đó chỉ được gửi dữ liệu sau khi nó nhận được phản hồi từ mạng. Nó cũng mong đợi gửi dữ liệu nằm chờ đó trong khi nó được thông tin về hiện trạng lưu lượng.

Mạng thực hiện nhiệm vụ điều khiển quản trị dựa trên những thông tin từ ứng dụng và độ sẵn sàng của các nguồn tài nguyên mạng, nó cũng xác nhận trả lời các thông số QoS của ứng dụng ngay khi lưu lượng vẫn còn nguyên hiện trạng đặc biệt, mạng điền đầy sự xác nhận của nó nhằm duy trì trạng thái lưu lượng và sau đó thực hiện phân chia gói, chính sách và xếp hàng thông minh dựa trên trạng thái đó.

Các loại hình dịch vụ tích hợp được cung cấp bao gồm:

+ *Giao thức dành sẵn tài nguyên (RSVP – Resource Reservation Protocol).* Có thể ứng dụng trong định tuyến tín hiệu.

+ *Cơ chế hàng đợi thông minh.* Có thể được sử dụng với RSVP cung cấp các loại hình dịch vụ sau:

+ *Dịch vụ tốc độ được đảm bảo (Guaranteed Rate Service).* Loại hình này cho phép dành sẵn độ rộng băng thông để phù hợp với những yêu cầu của chúng, ví dụ ứng dụng VoIP dành 8 kbit/s từ đầu cuối đến đầu cuối sử dụng loại hình dịch vụ này. Có thể sử dụng xếp hàng cân bằng trọng số (WFQ) kết hợp với giao thức dành sẵn tài nguyên (RSVP) để cung cấp loại hình dịch vụ này. Dịch vụ này được sử dụng cho các ứng dụng thời gian thực. Nó cung cấp sự đảm bảo về ngưỡng trễ và băng thông. Dịch vụ này đảm bảo rằng các gói sẽ đến đích trong một khoảng thời gian xác định và sẽ không bị loại bỏ khi hàng đợi bị

tràn miễn là lưu lượng đó tuân thủ đúng thoả thuận. Tuy nhiên dịch vụ này không điều khiển độ trễ nhỏ nhất và độ trễ trung bình. Do đó nó không điều khiển được *biến động trễ*. Nó chỉ đảm bảo độ trễ không vượt quá ngưỡng lớn nhất.

+ Dịch vụ tải được điều khiển (Controlled Load Service). Loại hình này cho phép các ứng dụng có độ trễ thấp và tốc độ lưu lượng cao thậm chí ngay cả khi tắc nghẽn. Ví dụ các ứng dụng thời gian thực (Real -Time) như khi phát lại băng ghi âm cuộc hội thoại có thể sử dụng loại hình dịch vụ này. Sử dụng RSVP kết hợp với *Weighted Random Early Detect* (WRED) cung cấp loại hình dịch vụ này. Là dịch vụ gần giống với dịch vụ cố gắng tối đa nhưng chỉ đối với các trường hợp mạng ít tải hoặc không tải. Tức là nó cố gắng đảm bảo rằng tất cả các gói sẽ được truyền đến đúng địa chỉ và độ trễ sẽ không vượt quá ngưỡng quy định. Tuy nhiên điều này còn phụ thuộc vào tình trạng mạng.

Động lực thúc đẩy mô hình *IntServ* chủ yếu do những lý do cơ bản sau đây:

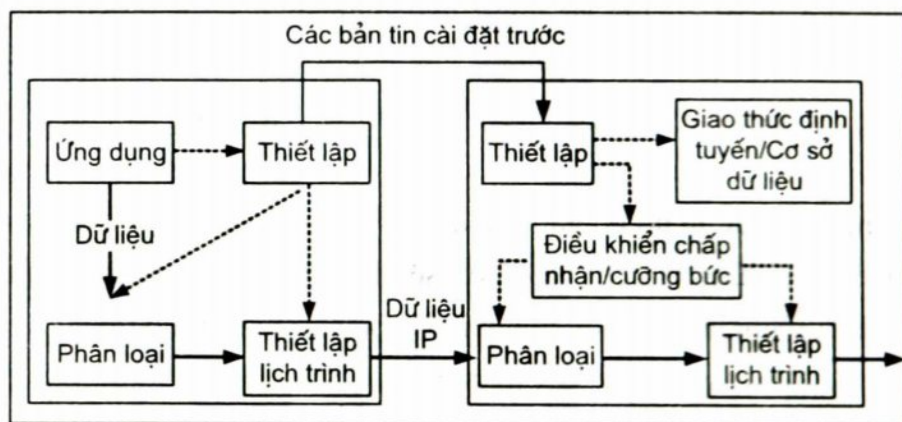
- Dịch vụ nỗ lực tối đa không còn đủ tốt nữa: ngày càng có nhiều ứng dụng khác nhau có những yêu cầu khác nhau về đặc tính lưu lượng được triển khai, đồng thời người sử dụng ngày càng yêu cầu cao hơn về chất lượng dịch vụ.

- Các ứng dụng đa phương tiện cả gói ngày càng xuất hiện nhiều: mạng IP phải có khả năng hỗ trợ không chỉ đơn dịch vụ mà phải hỗ trợ tích hợp đa dịch vụ của nhiều loại lưu lượng khác nhau từ thoại, số liệu đến video.

- Tối ưu hóa hiệu suất sử dụng mạng và tài nguyên mạng: đảm bảo hiệu quả sử dụng và đầu tư. Tài nguyên mạng sẽ được dự trữ cho lưu lượng có độ ưu tiên cao hơn, phần còn lại sẽ dành cho số liệu nỗ lực tối đa.

- Cung cấp dịch vụ tốt nhất: mô hình dịch vụ IntServ cho phép nhà cung cấp mạng cung cấp được dịch vụ tốt nhất khác biệt với các nhà cung cấp cạnh tranh khác.

Hình 6.19 là mô hình dịch vụ tích hợp (IntServ).



Hình 6.19: Mô hình dịch vụ tích hợp (IntServ)

Trong mô hình này có một số thành phần tham gia như sau:

- Giao thức thiết lập: Setup (Thiết lập) cho phép các máy chủ và các bộ định tuyến dự trữ động tài nguyên trong mạng để xử lý các yêu cầu của các luồng lưu lượng riêng, RSVP, Q.2931 là một trong những giao thức đó.

- Đặc tính luồng: xác định chất lượng dịch vụ QoS sẽ cung cấp cho luồng riêng biệt. Luồng được định nghĩa như một luồng

các gói từ nguồn đến đích có cùng yêu cầu về QoS. Về nguyên tắc có thể hiểu đặc tính luồng như băng tần tối thiểu mà mạng bắt buộc phải cung cấp để đảm bảo QoS cho luồng yêu cầu.

- Điều khiển lưu lượng: trong các thiết bị mạng (máy chủ, bộ định tuyến, chuyển mạch) có thành phần điều khiển và quản lý tài nguyên mạng cần thiết để hỗ trợ QoS theo yêu cầu. Các thành phần điều khiển lưu lượng này có thể được khai báo bởi giao thức báo hiệu như RSVP hay nhân công. Thành phần điều khiển lưu lượng bao gồm:

- Điều khiển chấp nhận: xác định thiết bị mạng có khả năng hỗ trợ QoS theo yêu cầu hay không;

- Thiết bị phân loại (Classifier): nhận dạng và lựa chọn lớp dịch vụ dựa trên nội dung của một số trường nhất định trong mào đầu gói;

- Thiết bị lập lịch (Scheduler): cung cấp các mức chất lượng dịch vụ QoS trên kênh ra của thiết bị mạng.

Các mức chất lượng dịch vụ cung cấp bởi IntServ bao gồm:

- Dịch vụ bảo đảm (GS): băng tần dành riêng, trễ có giới hạn và không bị thất thoát gói tin trong hàng. Các ứng dụng cung cấp thuộc loại này có thể kể đến: hội nghị truyền hình chất lượng cao, thanh toán tài chính thời gian thực,...

- Dịch vụ kiểm soát tải (CL): không đảm bảo về băng tần hay trễ nhưng khác nỗ lực tối đa ở điểm không giảm chất lượng một cách đáng kể khi tải mạng tăng lên. Phù hợp cho các ứng dụng không nhạy cảm lắm với độ trễ hay mất gói như truyền đa hướng audio/video chất lượng trung bình.

- Dịch vụ cố gắng tối đa.

Mô hình dịch vụ của *IntServ* được chuẩn hóa theo độ trễ, tỉ lệ mất gói và tốc độ từ đầu cuối đến đầu cuối. Tuy nhiên các mô hình này được phân biệt với nhau dựa trên nguyên tắc: "Các dịch vụ khác biệt dựa vào việc đặt chỗ tài nguyên trước cho các luồng của dịch vụ đó". Một số đặc điểm của kiến trúc *IntServ*:

Thứ nhất, *IntServ* là kiến trúc tập trung vào từng luồng (flow-centric), nghĩa là các cơ chế của bộ định tuyến được áp dụng vào các luồng riêng rẽ trong cùng một phiên ứng dụng. Các cơ chế của bộ định tuyến được áp dụng trên các gói của một luồng bao gồm:

- *Phân loại luồng:* để xác định gói đó thuộc về luồng nào.

- *Lập lịch gói tin:* thực hiện cấp phát tài nguyên. Nó quyết định gói nào gửi kế tiếp khi tuyến kết nối đi đã sẵn sàng.

- *Quản lý bộ đệm:* để phân bổ số bộ đệm cho luồng

- *Điều tiết tốc độ luồng hoặc loại bỏ tuân theo chính sách:* để điều khiển các đặc tính của luồng (ví dụ số cụm dữ liệu tối đa trong luồng).

Thứ hai, *IntServ* là một kiến trúc đầu cuối - đầu cuối, do đó nó yêu cầu sự kết hợp giữa các nhà cung cấp mạng khác nhau để có được sự đảm bảo về dịch vụ trên toàn bộ đường truyền từ đầu cuối - đầu cuối. *Internet* là một mạng có cấu trúc phân cấp, do đó việc thiết lập các thỏa thuận từ nhiều phía là rất khó khăn.

Thứ ba, *IntServ* được dựa trên việc dành trước tài nguyên, nghĩa là tài nguyên (băng thông và bộ đệm) phải được dành sẵn cho luồng trước khi phiên làm việc bắt đầu và kéo dài trong suốt

phiên làm việc. Để thực hiện điều này phải dùng giao thức báo hiệu. Giao thức báo hiệu truyền các yêu cầu về *QoS* của ứng dụng và các đặc tính của lưu lượng từ đầu cuối đến các bộ định tuyến dọc theo tuyến truyền dẫn. Nếu tài nguyên không đủ để đặt trước, báo hiệu sẽ thông báo cho đầu cuối biết luồng sẽ không được chấp nhận. Quá trình này gọi là *điều khiển quyền truy nhập (Admission Control)*. Giao thức báo hiệu được dùng trong kiến trúc *IntServ* là *RSVP*.

Mặc dù *IntServ* có khả năng cung cấp *QoS* cho các ứng dụng với mức độ đảm bảo nghiêm ngặt song sự phức tạp và khó khăn trong quản lý mạng cũng là nhược điểm của *IntServ*. Sự phức tạp thể hiện trong việc *IntServ* đòi hỏi các bộ định tuyến phải duy trì, xử lý dữ liệu và điều khiển tất cả các luồng. Các đường kết nối tốc độ Gigabit hay Terabit mang hàng triệu luồng đồng thời và điều này làm cho các bộ định tuyến trở nên quá tải. Các bộ định tuyến thế hệ sau cũng có thể xử lý hàng triệu luồng bằng cách lưu trữ các trạng thái từng luồng tại bộ định tuyến *biên*. Khó khăn trong việc quản lý do kiến trúc *IntServ* rất khó cài đặt, bảo trì và hoạt động. Ngoài ra kiến trúc yêu cầu các giao diện chương trình ứng dụng tại các đầu cuối cho các ứng dụng đa phương tiện.

IntsServ và RSVP

Mô hình các dịch vụ tích hợp sẽ cung cấp phương pháp để ứng dụng các mức dịch vụ cho luồng dữ liệu của "dịch vụ phân phối". IntsServ được mô tả như một cấu trúc gồm 2 bộ phận chính:

- *Bộ định tuyến (Router) và máy trạm (host)* có thể hiểu và có thể phân biệt được những mức độ dịch vụ khác nhau. Những

thiết bị này phải có khả năng xử lý các gói có đặc điểm nhất định (định tuyến đặc biệt, trật tự ưu tiên....). Chức năng này đôi khi được hiểu như là chức năng điều khiển của Intserv. Những ví dụ về chức năng điều khiển QoS bao gồm tải được điều khiển và độ tin cậy, điều đó được chỉ ra trong RFC 2211 và 2212.

- Dịch vụ được điều khiển sử dụng chức năng điều khiển đầu vào để xác định và ưu tiên những gói phù hợp với đặc điểm nhất định, làm cho lưu lượng đó có những điều kiện tương tự với những điều kiện trên mạng không quá tải. Khi lưu lượng đã được điều khiển, dịch vụ mạng dường như là "tốt nhất" nếu nói về độ trễ và độ tin cậy,... đó là những vấn đề không dự đoán trước được. Khi việc điều khiển QoS đã được đảm bảo, các phần tử mạng chấp nhận cung cấp một dịch vụ với độ trễ có khả năng dự đoán được về mặt toán học.

Mô hình Intserv phụ thuộc vào tác động của hai chức năng trên và những ứng dụng của chúng đối với luồng lưu lượng, một chuỗi các gói có liên quan với nhau theo một số cách và do đó có thể được xem như là một đơn vị. Ví dụ, một trạm trên Internet có thể thiết lập một kết nối với một trạm khác với mục đích truyền cuộc gọi VoIP, và tất cả các gói với những đặc tính của một cuộc gọi đặc biệt như thế có thể được xem như là một luồng đơn. Điều này có được là do các gói này có cùng chung một nơi đến và những địa chỉ IP nguồn, cùng cổng lớp truyền tải và những đặc tính khác. Đồng thời nếu 2 trạm cũng truyền tệp có lưu lượng thời gian không thực, các gói đó sẽ không cùng một luồng (ít nhất là khác nhau về số cổng). Do đó, các nút mạng có thể được phân biệt lưu lượng truyền tệp với lưu lượng VoIP.

Các luồng lưu lượng cần thiết cho hoạt động của RSVP và Intserv nói chung. Vì khung IntsServ sử dụng những đặc điểm của luồng lưu lượng như một phương tiện xác định lưu lượng cho định tuyến (chức năng điều khiển thiết lập), tùy theo những tham số kiểm soát QoS cụ thể (chức năng điều khiển QoS). Các bước cơ bản của Intserv bao gồm:

1. Ứng dụng gửi đi cho biết quá trình RSVP của nó về lưu lượng tạo ra cùng với những đặc tính của dòng lưu lượng. Ứng dụng cung cấp cho RSVP một bản mô tả về những yêu cầu mà một dòng lưu lượng phải có.

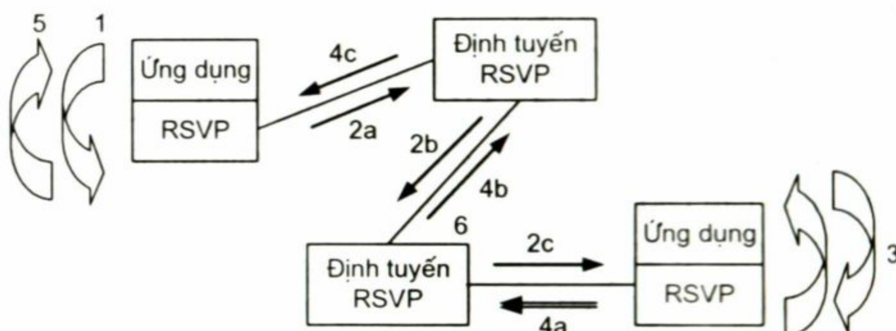
2. Một bản tin về cơ cấu RSVP chứa đựng hai nhân tố được chuyển đi theo đường dẫn tới người nhận đã định và được sửa đổi về thông tin và khả năng của mỗi nút khi nó được tiếp âm từng chặng một, tạo ra một bản mô tả về đường dẫn. Ví dụ, một phần những yêu cầu của dòng lưu lượng có thể trở thành đơn vị có thể truyền dẫn tối đa (MTU) có kích cỡ 1 KB và mỗi bộ định tuyến dọc theo đường dẫn sẽ cho thấy khả năng của nó (tính sẵn sàng của nó để cung cấp MTU đó cho dòng lưu lượng).

3. Khi bản tin setup đến được người nhận, RSVP sẽ mô tả đường dẫn cho ứng dụng, ứng dụng này có thể sử dụng đường dẫn đó vào việc lựa chọn những thuộc tính hợp lý để yêu cầu. Khi ứng dụng đã cho ta được yêu cầu của nó, nó sẽ trao trả lại yêu cầu cho RSVP.

4. RSVP chuyển chuỗi yêu cầu đó dọc theo đường dẫn đến người gửi, đòi hỏi những thuộc tính đó từ các bộ định tuyến.

5. Khi RSVP của người gửi nhận được bản tin, nó sẽ thông báo cho ứng dụng gửi về sự tồn tại của đường dẫn cùng với những đặc tính đặc biệt thuận tiện cho việc sử dụng.

6. Các bộ định tuyến thực hiện "sự duy trì" bằng cách làm cho các đặc tính của dòng lưu lượng đã được đề cập ở trong bản tin setup của người gửi phù hợp với những thuộc tính chỉ ra trong yêu cầu của người nhận.



Hình 6.20: Quá trình thiết lập RSVP

Trong khi quá trình IntsServ là tương đối phức tạp, nó rất xứng đáng với cố gắng để duy trì một tập hợp những thuộc tính của mạng cho một lưu lượng cụ thể - ít nhất là từ quan điểm của ứng dụng cá nhân. Sẽ không khó gì trong việc tưởng tượng ra mọi thiết bị trên Internet có thể yêu cầu mạng một mức độ dịch vụ cần thiết cho những gì mà thiết bị vận hành ở một thời điểm cụ thể hoặc không có một yêu cầu gì cả khi chỉ yêu cầu dịch vụ "cố gắng tối đa".

Tuy nhiên mô hình IntsServ có những hạn chế nhất định, điều này có liên quan đến quy mô ngày càng mở rộng nhanh chóng của Internet hơn là có liên quan đến thiết kế giao thức. Trong khi có thể không có khó khăn gì cho một máy chủ truyền tải hàng trăm hoặc hàng ngàn những yêu cầu của RSVP mà nó phải đảm nhiệm thay mặt khách hàng đồng thời với việc kết nối

tới nó, những bộ định tuyến ở trung tâm của mạng có thể sẽ có những thông tin về tình trạng của luồng lưu lượng đi qua có liên quan đến không phải là một máy chủ và những khách hàng của nó mà là nhiều máy chủ và tất cả các khách hàng của chúng. Một số bộ định tuyến tại vị trí có nguồn lưu lượng cao đôi khi cần thiết phải duy trì trạng thái hàng triệu cuộc hội thoại cùng một lúc.

RSVP có nhược điểm của tất cả những giao thức trong những nút khi mà bản thân lưu lượng không tồn tại được, đòi hỏi nút phải cung cấp năng lượng vào bộ nhớ để dự đoán được lưu lượng và để duy trì thông tin về tín hiệu đó. Những bộ định tuyến đường trục hoạt động để định tuyến các gói và duy trì thông tin sử dụng giao thức định tuyến, mà không cần phải chịu thêm chi phí gia tăng của việc xử lý thông tin về trạng thái. Việc này sẽ chỉ tăng độ trễ trong định tuyến. Khi mạng Internet tăng về quy mô và tốc độ, số lượng trạng thái lưu lượng đòi hỏi việc duy trì sẽ tăng lên.

Một hạn chế khác đối với việc triển khai RSVP ở diện rộng đó là sự phối hợp của những nhà cung cấp nội bộ để những dòng RSVP được đăng ký, tìm kiếm và hoạt động dọc theo đường dẫn những nhà cung cấp. Khi yêu cầu của nhà cung cấp hỗ trợ chức năng điều khiển QoS giống nhau, những mức độ dịch vụ có khả năng thích hợp cho những dòng nói trên và công suất đủ cho số luồng nói trên sẽ hạn chế khả năng hợp tác của những nhà cung cấp dịch vụ Internet.

Do những hạn chế đó IntsServ và RSVP được nhiều người xem là chỉ có ích ở biên đường dẫn, cho những bộ phận riêng rẽ

của cuộc gọi VoIP đi qua mạng của nhà cung cấp đó và cho những mạng của nhà cung cấp riêng rẽ dành cho VoIP và những dịch vụ thời gian thực khác, chừng nào chúng còn trực thuộc một cơ quan quản lý đơn nhất. Tuy nhiên. IntsServ rõ ràng không phải là câu trả lời cho sự hỗ trợ VoIP ở diện rộng hoặc những dịch vụ thời gian thực khác trên Internet nói chung.

6.4.2.3 Mô hình phân quyền ưu tiên – cơ chế cung cấp dịch vụ khác biệt (DS - Differentiated Service)

Differentiated Service hay viết tắt là *DiffServ* là một mô hình đa dịch vụ, chúng có thể làm thỏa mãn các yêu cầu QoS khác nhau. Tuy nhiên, nó không giống mô hình dịch vụ *IntServ*. Một ứng dụng sử dụng mô hình dịch vụ *DiffServ* không cần xác định rõ ràng tín hiệu tới bộ định tuyến trước khi gửi dữ liệu.

Cơ chế cung cấp dịch vụ khác biệt được định nghĩa bởi IETF. Để cung cấp các lớp dịch vụ khác biệt trên đường truyền Internet, hỗ trợ nhiều loại ứng dụng và đáp ứng nhu cầu của các doanh nghiệp. Cơ chế này hoạt động dựa trên việc cung cấp chất lượng dịch vụ cho từng nhân tố sẽ tạo nên các dịch vụ khác nhau hơn là cho bản thân các dịch vụ. Cách thức truyền thông trên mạng cũng như các tài nguyên mạng sẽ được phân loại dựa trên các tiêu chuẩn quản lý dải thông. Để đảm bảo chất lượng dịch vụ, việc phân loại trên phải đưa ra những cơ chế ưu tiên cho những ứng dụng có nhu cầu cao hơn.

Trong mô hình dịch vụ *DiffServ*, mạng cố gắng phân loại dịch vụ dựa trên lý thuyết QoS cho mỗi gói tin. Đặc điểm kỹ thuật này có thể thực hiện bằng nhiều cách khác nhau, ví dụ sử

dụng trường bit IP Precedence trong tiêu đề IP hoặc địa chỉ nguồn và đích, mạng sử dụng đặc điểm kỹ thuật QoS để phân loại, đánh dấu, nhận dạng và chính sách lưu lượng nhằm xếp hàng đợi thông minh.

DiffServ định nghĩa lại 6 trong số 8 bit trong trường ToS của phần mào đầu trong gói IP cho phép các bit ToS được sử dụng để phân biệt các ứng dụng. 6 bit này tổ hợp ra 64 lớp dịch vụ, nó đại diện cho các loại ứng dụng khác nhau và sẽ được chuẩn hóa giữa tất cả các ISP và các bộ định tuyến. Chuẩn *DiffServ* rất hấp dẫn, nhưng tất nhiên là tất cả các bộ định tuyến phải hiểu và tuân theo các loại QoS của *DiffServ*. *DiffServ* không có các đảm bảo thực hiện QoS hoàn toàn. Ví dụ, *DiffServ* tốt nhất có thể làm cho VoIP là đảm bảo rằng các gói thoại được xếp hàng đầu tiên tới cổng ra.

Mô hình dịch vụ *DiffServ* được sử dụng dành cho những ứng dụng đặc biệt và để cung cấp đầu cuối tới đầu cuối QoS. Tiêu biểu cho loại hình dịch vụ này phù hợp cho việc tập trung lưu lượng, bởi vì nó thực hiện việc phân loại lưu lượng ở cấp độ thô.

Các mô hình cung cấp dịch vụ *DiffServ* như sau:

- Tốc độ truy nhập cam kết (CAR - Committed Access Rate), loại hình này thực hiện phân loại gói thông qua IP Precedence và thiết lập nhóm QoS. CAR thực hiện đo và chính sách của lưu lượng, quản lý độ rộng băng thông.

- Xếp hàng đợi thông minh như WRED (phát hiện sớm ngẫu nhiên theo trọng số) và WFQ (xếp hàng theo trọng số) là các phương pháp để tránh tắc nghẽn và những đặc tính tương đương

dựa trên bộ xử lý giao diện đa năng (VIP – Versatile Interface Processor) gọi là VIP – Distributed WRED và VIP – Distributed WFQ. Những đặc điểm này có thể được sử dụng với tốc độ truy nhập cam kết (CAR) để phân phối các dịch vụ *DiffServ*.

Mô hình dịch vụ khác biệt cho phép các nhà cung cấp dịch vụ đưa ra cho mỗi khách hàng đồng thời nhiều dịch vụ. Các dịch vụ này khác nhau cơ bản về hiệu suất, kèm theo đó là chi phí cho dịch vụ đó.

Thoả thuận giữa nhà cung cấp dịch vụ và người sử dụng được thực hiện tại biên của mạng. Các thoả thuận này được gọi là các *thoả thuận về mức dịch vụ (SLA)*. Một phần của *SLA* sẽ đặc tả mức dịch vụ trên phương diện kỹ thuật và được gọi là *SLS (Service Level Specification)*. *SLS* có thể là tĩnh hoặc động. *SLS* tĩnh được cài đặt bởi người dùng hoặc nhà cung cấp. Các đặc tả này có thể chỉ ra những thay đổi về mức dịch vụ tại một thời điểm nhất định trong ngày, tuần hoặc tháng. Ngược lại, *SLS* động có thể thay đổi thường xuyên. Sự thay đổi này có thể là do sự thay đổi của tải lưu lượng so với ngưỡng hoặc thay đổi do giá cả của nhà cung cấp đưa ra khi tải lưu lượng thay đổi. *SLS* động thay đổi không cần sự can thiệp của con người. Tuy nhiên nó được quản lý bởi một giao thức gọi là *Bandwidth broker* (môi giới băng thông) được cài đặt trong miền *DiffServ*.

Ban đầu, việc đưa ra mô hình IntServ đã có vẻ như giải quyết được nhiều vấn đề liên quan đến QoS trong mạng IP. Tuy nhiên trên thực tế, mô hình này không thực sự đảm bảo được QoS xuyên suốt (end-to-end). Đã có nhiều cố gắng để thay đổi

điều này nhằm đạt được một mức QoS cao hơn cho mạng IP và một trong những cố gắng đó là sự ra đời của *DiffServ*. *DiffServ* sử dụng việc đánh dấu gói và xếp hàng theo loại để hỗ trợ các dịch vụ ưu tiên qua mạng IP. Hiện tại IETF đã có một nhóm làm việc *DiffServ* để đưa ra các tiêu chuẩn RFC về *DiffServ*.

Nguyên tắc cơ bản của *DiffServ* như sau:

- Định nghĩa một số lượng nhỏ các lớp dịch vụ hay mức ưu tiên. Một lớp dịch vụ có thể liên quan đến đặc tính lưu lượng (băng tần min- max, kích cỡ burst, thời gian kéo dài burst..).

- Phân loại và đánh dấu các gói riêng biệt tại biên của mạng vào các lớp dịch vụ.

- Các thiết bị chuyển mạch, bộ định tuyến trong mạng lõi sẽ phục vụ các gói theo nội dung của các bít đã được đánh dấu trong mào đầu của gói.

- Với nguyên tắc này, *DiffServ* có nhiều lợi thế hơn so với *IntServ:*

- Không yêu cầu báo hiệu cho từng luồng;

- Dịch vụ ưu tiên có thể áp dụng cho một số luồng riêng biệt cùng một lớp dịch vụ. Điều này cho phép nhà cung cấp dịch vụ dễ dàng cung cấp một số lượng nhỏ các mức dịch vụ khác nhau cho khách hàng có nhu cầu;

- Không yêu cầu thay đổi tại các máy chủ hay các ứng dụng để hỗ trợ dịch vụ ưu tiên. Đây là công việc của thiết bị biên;

- Hỗ trợ rất tốt dịch vụ VPN.

- Tuy nhiên có thể nhận thấy *DiffServ* cần vượt qua một số vấn đề như:

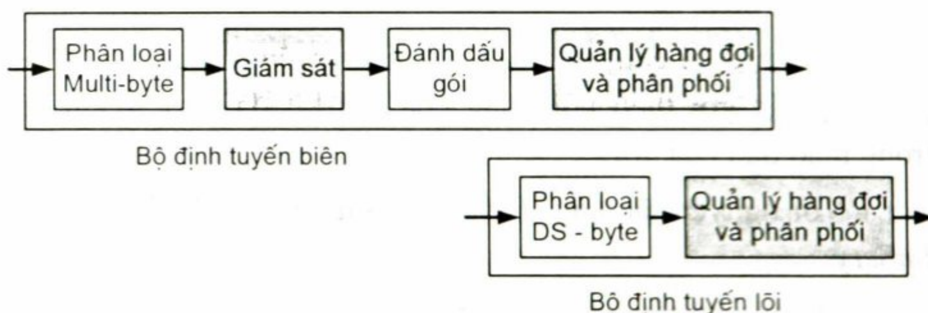
- Không có khả năng cung cấp băng tần và độ trễ đảm bảo như GS của IntServ hay ATM;

- Thiết bị biên vẫn yêu cầu Bộ Phân loại (Classifier) chất lượng cao cho từng gói giống như trong mô hình IntServ;

- Vấn đề quản lý trạng thái bộ phân loại của một số lượng lớn các thiết bị biên là một vấn đề không nhỏ cần quan tâm;

- Chính sách khuyến khích khách hàng trên cơ sở giá cước cho dịch vụ cung cấp cũng ảnh hưởng đến giá trị của *DiffServ*.

Mô hình của *DiffServ* tại biên và lõi được mô tả trong hình 6.21 dưới đây:



Hình 6.21: Mô hình DiffServ tại biên và lõi của mạng

Mô hình *DiffServ* bao gồm một số thành phần như sau:

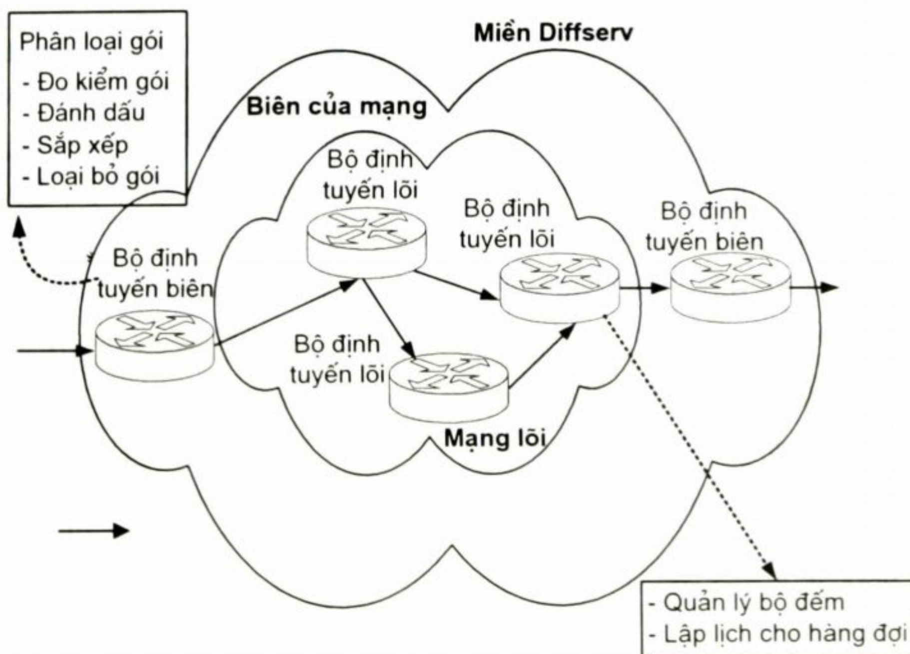
- DS-Byte: byte xác định *DiffServ* là thành phần ToS của IPv4 và trường loại lưu lượng IPv6. Các bit trong byte này thông báo gói tin được mong đợi nhận được thuộc dịch vụ nào.

- Các thiết bị biên (Bộ định tuyến biên): nằm tại lối vào hay lối ra của mạng cung cấp *DiffServ*.

- Các thiết bị bên trong mạng *DiffServ*.

- Giám sát: các công cụ và nhà quản trị mạng giám sát và đo kiểm đảm bảo SLA giữa mạng và người dùng.

DiffServ phân biệt rất rõ ràng các bộ định tuyến nằm tại biên của mạng (tức là các bộ định tuyến đầu vào và đầu ra) với các bộ định tuyến lõi để có thể đáp ứng được yêu cầu mở rộng mạng. Nó loại bỏ rất nhiều chức năng ra khỏi bộ định tuyến lõi của mạng và đặt chúng vào trong bộ định tuyến biên. Hình 6.22 mô tả cơ chế bộ định tuyến trong miền *DiffServ*:



Hình 6.22: Chức năng của bộ định tuyến biên và lõi trong miền DiffServ

* Bộ định tuyến biên

Các bộ định tuyến được đặt tại biên của miền được gọi là đầu vào hoặc đầu ra tuỳ theo các hướng khác nhau của lưu

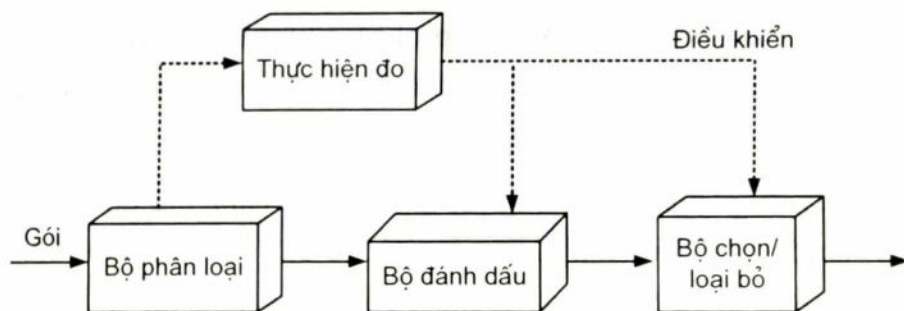
lượng. Lưu lượng đi vào miền tại một bộ định tuyến đầu vào và rời miền tại bộ định tuyến đầu ra.

Các bộ định tuyến đầu vào phải thực hiện phân loại lưu lượng *(traffic classification)* và điều khiển lưu lượng *(traffic conditioning)*. Phân loại lưu lượng nhận dạng các đặc tính của một tập các lưu lượng, các lưu lượng này sau đó có thể được nhận một dịch vụ từ bộ điều khiển lưu lượng *(traffic conditioner)* và được xếp vào một lớp dịch vụ. Phân loại lưu lượng có thể thực hiện trên một luồng riêng biệt hoặc một luồng gộp. Điều khiển lưu lượng thực hiện đo *(metering)*, đánh dấu *(marking)*, điều tiết *(sharping)*, và loại bỏ *(droping)*. Chức năng đo tiến hành đo các luồng đã được chấp nhận và kiểm tra liệu luồng có thỏa mãn hồ sơ lưu lượng *(traffic profile)* hay không. Chức năng này cung cấp đầu vào cho 3 chức năng tiếp theo. Ví dụ về kiểm tra hồ sơ lưu lượng của *tocken-bucket:*

```
if Codepoint = X
then
use tocken-bucket r,b
if tocken available
then
    forward packet
else
    discard packet
end if
end if
```

Hồ sơ này chỉ ra rằng tất cả các gói được đánh dấu với *codepoint* (điểm mã) là *X* sẽ được đo bằng bộ đo *token-bucket* với tốc độ r và kích thước bucket là b. Nếu còn thẻ trong giỏ thì gói sẽ được coi là hợp lệ và cho đi qua, ngược lại gói sẽ bị loại.

Các chức năng của bộ định tuyến biên được thể hiện trong hình vẽ 6.23:



Hình 6.23: Cơ chế của bộ định tuyến biên

Chức năng đánh dấu thiết lập cho trường *DS* của gói một giá trị *DS codepoint* tương ứng với một *PHB* hoặc để bảo đảm rằng chỉ các *codepoint* hợp lệ mới được sử dụng trong miền. Chức năng điều tiết kiểm soát gói bằng cách làm trễ lưu lượng do đó tốc độ không bị vượt quá so với tốc độ được chỉ ra trong hồ sơ. Chức năng loại bỏ gói thực hiện loại bỏ gói khi nó không thỏa mãn hồ sơ lưu lượng.

* *Bộ định tuyến lõi*

Các bộ định tuyến lõi trong miền *DiffServ* phải đảm bảo rằng các yêu cầu dịch vụ được chỉ ra bởi *SLS* phải được đáp ứng. Hai cơ chế chính được sử dụng trong bộ định tuyến lõi để thực hiện đảm bảo dịch vụ cho các luồng gộp là xếp hàng (*buffering)* và chương trình đọc gói *(packet scheduling*).

Xếp hàng quyết định xem bao nhiêu dung lượng của bộ đệm được dành cho các loại dịch vụ và bao nhiêu gói bị loại bỏ trong trường hợp có tắc nghẽn. Chức năng này căn cứ vào ngưỡng loại bỏ *(threshold droping)* để loại gói. Các gói từ các lớp khác nhau được xếp hàng vào trong một bộ đệm và được phục vụ theo cơ chế FIFO. Mỗi lớp trong dịch vụ i sẽ có ngưỡng loại bỏ là B_i và một gói của lớp i chỉ được chấp nhận nếu dung lượng bộ đệm dành cho lớp đó là $B(t) \le B_i$, ngược lại, gói sẽ bị loại bỏ với xác suất là p_i.

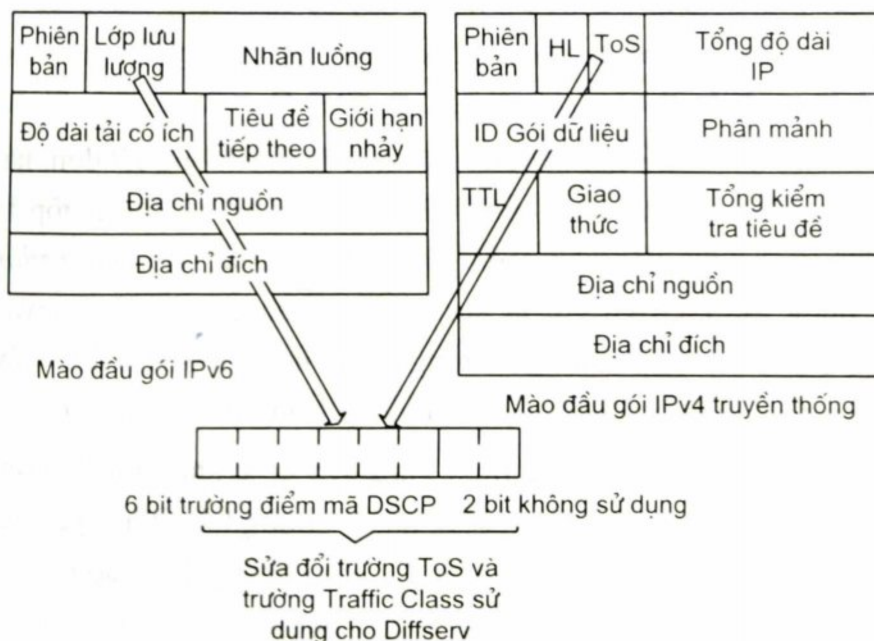
Chương trình đọc gói quyết định thứ tự hàng đợi được phục vụ, do đó yêu cầu băng thông và độ trễ của các loại lưu lượng khác nhau được đáp ứng. Trong chức năng này, các gói thuộc các lớp khác nhau được gửi tới các hàng đợi khác nhau và các hàng đợi được phục vụ như thế nào tùy thuộc vào thuật toán được sử dụng để đọc gói ra.

Một điểm nổi bật trong mạng *DiffServ* là không có điều khiển quyền truy nhập ở bộ định tuyến lõi, do đó rất khó để duy trì các yêu cầu dịch vụ của các lớp một cách định lượng. Bản thân mạng *DiffServ* không thể một mình giải quyết vấn đề này. Một số cơ chế khác như định tuyến ràng buộc/kiểm soát lưu lượng *(constrained-based routing/traffic engineering)*, RSVP và môi giới băng thông *(bandwidth brocker)* được sử dụng để quản lý và tối ưu hóa việc sử dụng tài nguyên trong miền *DiffServ*.

Trường ToS và trường Traffic class trong DiffServ

Khi xem xét giải pháp cho QoS của Internet giống như Intserv và RSVP tại ISP và vùng định tuyến, người ta chú ý nhiều

đến octet thứ 2 ít dùng đến trong mào đầu IP, đó là trường ToS. Vào tháng 12 năm 1997, nhóm làm việc về dịch vụ tích hợp của IETF đã làm các thành viên nhóm làm việc *DiffServ* "chao đảo" khi có ý định là sáng tạo ra một phương pháp có quy mô về "các dịch vụ khác biệt". Ngược lại với giải pháp Intserv và RSVP, phương pháp *DiffServ* cho phép các bộ định tuyến xác định được trạng thái truyền tải của mọi gói dựa trên thành phần của một trường điểm mã DS (DSCP) bằng cách sử dụng trường *ToS* trong tiêu đề gói IPv4 hoặc trường *Traffic class* trong tiêu đề gói IPv6 và đổi tên nó thành trường DS như hình 6.24. Trường này được sử dụng để mang thông tin về các yêu cầu dịch vụ.



Hình 6.24: DiffServ với trường ToS và trường Traffics Class

Trường ToS bao gồm 2 trường con: 3 bit ưu tiên để cung cấp 8 mức ưu tiên và 3 hoặc 4 bit thuộc tính. Số lượng chính xác của bit ở trong trường con tùy thuộc vào thời gian thực hiện. Các octet *DiffServ* này bao gồm 01 trường DS 6 bit, 2 bit khác không sử dụng. 6 bit được dùng để lựa chọn ra điểm mã dịch vụ khác biệt (DSCP-*DiffServ Code Point*). Việc sử dụng trường đó hoàn toàn không tương thích với việc sử dụng trường "truyền thống" đã được chỉ ra trong RFC 791. Hầu hết mọi người không nhìn thấy nó như một vấn đề dẫn đến sử dụng không đồng bộ hiện nay.

Khách hàng yêu cầu một mức hiệu suất cụ thể trên mỗi gói bằng cách điền vào trường *DS* của mỗi gói một giá trị cụ thể gọi là *DS Code Point (DSCP)*. Giá trị này sẽ cho các bộ định tuyến biết phải xử lý gói như thế nào và mỗi cách xử lý ứng với một *DSCP* gọi là hành vi xử lý mỗi chặng *(Per-Hop Behavior - PHB)*.

Một đặc điểm nổi bật của *DiffServ* là các luồng *IP* đơn, hay còn gọi là *microflow* có cùng yêu cầu về *QoS* sẽ được gộp lại thành một luồng tổng hay còn gọi là *macroflow*. Mỗi *macroflow* sử dụng lớp dịch vụ gọi là tập hợp hoạt động xử lý *(Behavior Aggregate - BA)* tương đương với một *PHB*. Như vậy, *macroflow* là đơn vị nhỏ nhất của luồng trong mô hình *DiffServ*.

Mô hình *DiffServ* xác định vùng DS có chứa một số lượng lớn các bộ định tuyến kết nối lẫn nhau, những bộ định tuyến này hiểu được những khái niệm tương tự cho những hoạt động từng bước (PHB). Mô hình này giả thiết vùng là một mạng của nhà cung cấp một dịch vụ đơn nhất hoặc một số mạng khác dưới sự kiểm soát của một cơ quan quản lý duy nhất, thông qua cơ quan

này có thể xác định được PHB. Việc cho phép truy cập vào và ra khỏi DS là các bộ định tuyến được biết đến như những nút ranh giới DS, những nút này "phân loại" lưu lượng vào vùng dựa trên một đặc tính lưu lượng của các nhà quản lý mạng. Việc lưu lượng thích hợp với một đặc tính cụ thể có thể có một giá trị CP nhất định. Giá trị này chỉ ra rằng một PHB nhất định sẽ được kế tiếp theo bởi những bộ định tuyến ở trong vùng đó. Giá trị này có thể có hoặc không có ý nghĩa tương tự (hoặc bất kỳ) với những bộ định tuyến ở ngoài vùng DS đó.

Cấu trúc dịch vụ phân biệt tùy thuộc vào sự hiểu biết nói chung về những quy tắc PHB và về việc chúng có liên quan như thế nào đến một QoS cụ thể sẽ có được từ mạng. Để đạt được mục tiêu này, *DiffServ* cho rằng lưu lượng qua ranh giới vùng DS sẽ phải được tính đến trong những thỏa thuận. Ví dụ những thỏa thuận về mức độ dịch vụ (SLA) và những thỏa thuận về điều kiện lưu lượng (TCA). Trong khi một số PHB sẽ được tiêu chuẩn hóa bởi IANA và độ an toàn (chẳng hạn như PHB "mặc định" tương tự lưu lượng Internet "thường lệ"), những nhà cung cấp sẽ được tự do, thoải mái trong việc xác định những điểm mã của họ và PHB để đưa ra những mức độ dịch vụ của họ.

6.4.3. Đánh giá các mô hình

Mạng Internet hiện nay như chúng ta biết không được thiết kế để truyền các luồng thông tin đa phương tiện chất lượng cao và dữ liệu thời gian thực. Nó là mạng phi kết nối, không có sự đảm bảo về chất lượng và chia sẻ bình đẳng tài nguyên mạng đối với mọi người sử dụng. Do đó, mạng Internet hiện nay gọi là

mạng cố gắng tối đa (best effort) và không thể cung cấp QoS. Hiện nay IETF đang nghiên cứu hai cơ chế có thể hỗ trợ QoS trong mạng Internet, đó là mô hình dịch vụ tích hợp (*Integrated Services*) và dịch vụ khác biệt (*Differentiated Services*). Cả hai phương pháp này đã được xem xét và đánh giá. Trong đó phương pháp theo mô hình dịch vụ tích hợp khó làm việc tốt trong mạng lớn vì có quá nhiều lưu lượng tạo ra do sử dụng giao thức dành trước tài nguyên (RSVP) và các thành phần mạng phải lưu giữ các bảng dịch vụ và trạng thái lớn. Còn phương pháp mô hình dịch vụ khác biệt thích hợp thì mềm dẻo hơn với mạng lớn và cũng được ưa dùng hơn. Các gói dữ liệu được phân lớp và được gán vào một trong các lớp dịch vụ bởi bộ định tuyến đầu vào mạng qua cơ chế đánh dấu gói. Các bộ định tuyến lõi đơn giản chỉ phải chuyển các gói theo lớp dịch vụ của chúng.

Bảng 6.7. So sánh DiffServ và IntServ

	IntServ	**DiffServ**
Bản chất của việc phân chia dịch vụ	Phân chia theo từng luồng	Phân chia theo từng lớp
Cơ sở để phân loại lưu lượng	Một số trường trong tiêu đề	Trường *DS*
Giao thức báo hiệu	Cần thiết phải có (*RSVP*)	Không yêu cầu
Phạm vi phân chia dịch vụ	Trên toàn chặng từ đầu cuối đến đầu cuối	Trên từng chặng
Khả năng cung cấp	Bị giới hạn bởi số lượng luồng	Bị giới hạn bởi số lượng lớp
Quản lý mạng	Tương tự như mạng chuyển mạch kênh	Tương tự như mạng *IP* hiện nay

THUẬT NGỮ VÀ VIẾT TẮT

ARP	Addresss Resolution Protocol	Giao thức phân giải địa chỉ
AS	Autonomous System	Hệ thống tự trị
ATM	Asynchronous Transfer Mode	Phương thức truyền tải không đồng bộ
BA	Behavior Aggregate	Tập hợp hoạt động xử lý
BGP	Border Gateway Protocol	Giao thức định tuyến cổng miền.
CELP	Code Excited Linear Prediction	Dự báo tuyến tính kích thích mã
COPS	Common Open Policy Service	Dịch vụ chính sách mở chung
CIDR	Classless InterDomain Routing	Định tuyến liên vùng không phân lớp
CSRC	Contributing Source Identifier	Bộ nhận dạng tập hợp nguồn
CRC	Cyclic Redundancy Check	Kiểm tra vòng dư
CTI	Computer Telephony Integration	Tích hợp điện thoại máy tính
DHCP	Dynamic Host Configuration Protocol	Giao thức cấu hình trạm động
DLCI	Data Link Connection Identifier	Nhận dạng kết nối lớp liên kết dữ liệu
DNS	Domain Name System	Hệ thống tên vùng
DS	Differentiated Service	Dịch vụ khác biệt
EGP	Edge Gateway Protocol	Giao thức định tuyến cổng biên
ETSI	European Telecommunication Standard Institute	Viện Tiêu chuẩn Viễn thông châu Âu
FDDI	Fiber Distributed Data Interface	Giao diện dữ liệu phân bố quang

FTP	File Transfer Protocol	Giao thức truyền tệp
HDLC	High-level Data Link Control	Điều khiển liên kết dữ liệu mức cao
HTTP	Hyper Text Transfer Protocol	Giao thức truyền dữ liệu siêu văn bản
IANA	Internet Assigned Numbers Authority	Cơ quan chỉ định số Internet
ICMP	Internet Control Message Protocol	Giao thức bản tin điều khiển Internet
IETF	International Engineering Task Force	Tổ chức tiêu chuẩn kỹ thuật quốc tế cho Internet
IEEE	Institute of Electrical and Electronic Engineers	Viện kỹ thuật điện và điện tử
IGMP	Internet Group Management Protocol	Giao thức quản lý nhóm liên mạng
IGP	Interior Gateway Protocol	Giao thức cổng mạng nội bộ
IN	Intelligent Network	Mạng thông minh
IP	Internet Protocol	Giao thức Internet
IPv4	IP version 4	IP phiên bản 4
IPv6	IP version 6	IP phiên bản 6
ISDN	Integrated Service Digital Network	Mạng số liên kết đa dịch vụ
ISO	International Organization for Standardization	Tổ chức tiêu chuẩn hóa quốc tế
ISP	Internet Services Provider	Các nhà cung cấp dịch vụ Internet
IS-IS	Intermediate System – Intermediate System routing exchange protocol	Giao thức trao đổi định tuyến IS-IS
ITU-T	International Telecommunication Union – Telecommunication Standardization	Ban Tiêu chuẩn Viễn thông của ITU

LSA	Link State Advertisement	Quảng cáo trạng thái liên kết
MAC	Media Access Controller	Thiết bị điều khiển truy nhập môi trường
MCS	Multipoint Communications System	Hệ thống liên lạc đa điểm
MCU	Multipoint Control Unit	Khối điều khiển đa điểm
MG	Media Gateway	Cổng nối phương tiện
MGC	Media Gateway Controller	Thiết bị điều khiển cổng nối phương tiện
MGCP	Media Gateway Control Protocol	Giao thức điều khiển cổng nối phương tiện
MPE	MultiPulse Excite	Phương pháp kích thích đa xung
MPLS	MultiProtocol Label Switching	Chuyển mạch nhãn đa giao thức
NGN	Next Generation Network	Mạng thế hệ sau
NNI	Network-Network Interface	Giao diện mạng - mạng
OSP	Open Settlement Protocol	Giao thức thanh toán mở
OSPF	Open Shortest Path First	Giao thức định tuyến OSPF
PDU	Protocol Data Unit	Đơn vị dữ liệu giao thức
PE	Provider Edge	Thiết bị định tuyến biên phía nhà cung cấp
PHB	Per-Hop Forwarding Behaviour	Hành vi xử lý mỗi chặng
PPP	Point to Point Protocol	Giao thức điểm - điểm
PSTN	Public Switch Telephone Network	Mạng chuyển mạch thoại công cộng
PVC	Permanent Virtual Circuit	Kênh ảo cố định
QoS	Quality of Service	Chất lượng dịch vụ
RARP	Reverse Address Resolution Protocol	Giao thức phân giải địa chỉ ngược
RAS	Registration, Authentication and Status	Giao thức RAS (Đăng ký, chứng thực và trạng thái)

RFC	Request for Comment	Các tài liệu về do IETF đưa ra
RIP	Realtime Internet Protocol	Giao thức Internet thời gian thực
RSVP	Resource Reservation Protocol	Giao thức dành trước tài nguyên (hỗ trợ QoS)
RTCP	RealTime Control Protocol	Giao thức điều khiển thời gian thực
RTSP	Real Time Streaming Protocol	Giao thức kiểm soát luồng dữ liệu
RTP	RealTime Transport Protocol	Giao thức truyền tải thời gian thực
SAP	Session Advertisement Protocol	Giao thức quảng cáo trong phiên kết nối
SDP	Session Description Protocol	Giao thức mô tả các phiên kết nối đa phương tiện
SGF	Signalling Gateway Function	Khối chức năng cổng báo hiệu
SIP	Session Initiation Protocol	Giao thức khởi tạo phiên
SLA	Service Level Agreement	Thỏa thuận kết nối mức dịch vụ giữa nhà cung cấp và khách hàng
SLS	Service Level Specification	Đặc điểm kỹ thuật mức dịch vụ
SNAP	Service Node Access Point	Điểm truy nhập nút dịch vụ
SNI	Signalling Network Interface	Giao diện mạng báo hiệu
SNMP	Simple Network Management Protocol	Giao thức quản lý mạng đơn giản
SONET	Synchronous Optical Network	Mạng truyền dẫn quang đồng bộ
SP	Service Provider	Nhà cung cấp dịch vụ
SPF	Shortest Path First	Giao thức định tuyến đường ngắn nhất
SVC	Switched Virtual Circuit	Kênh ảo chuyển mạch
TCA	Traffic Control Agreement	Thỏa thuận về điều khiển lưu lượng
TCP	Transport Control Protocol	Giao thức điều khiển truyền tải

TE	Terminal Equipment	Thiết bị đầu cuối
TELNET	TErminaL NETwork	Đầu cuối đăng nhập mạng
TIPHON	Telecommunications and Internet Protocol Harmonization Over Networks	Cân đối giữa Viễn thông và giao thức Internet trên các mạng
UDP	User Data Protocol	Giao thức dữ liệu người sử dụng
UNI	User Network Interface	Giao diện mạng - người sử dụng
VAD	Voice Activity Detector	Bộ phát hiện khoảng lặng
VCI	Virtual Circuit Identifier	Trường nhận dạng kênh ảo trong tế bào
VLSM	Variable Length Subnet Masking	Mặt nạ mạng con chiều dài biến thiên
VoIP	Voice over IP	Thoại qua giao thức IP
VPI	Virtual Path Identifier	Nhận dạng đường ảo
VPN	Virtual Private Network	Mạng riêng ảo
VPNID	Virtual Private Network IDentifier	Nhận dạng mạng riêng ảo
VR	Virtual Router	Bộ định tuyến ảo
VSC	Virtual Switched Controller	Khối điều khiển chuyển mạch ảo
VSF	Virtual Switched Function	Khối chức năng chuyển mạch ảo
WAN	Wide Area Network	Mạng diện rộng
WFQ	Weighted Factor Queque	Hàng đợi theo trọng số

TÀI LIỆU THAM KHẢO

[1] Jonathan Davidson, "*Voice over IP Fundamentals*", Cisco Press, 2000.

[2] Walter Goralski and Matthew C. Kolon, *"IP Telephony"*, McGraw-Hill, 2000.

[3] Behrouz A.Forouzan, *"TCP/IP Protocol Suite"*, McGraw-Hill, 2000.

[4] David J. Wright, *"Voice over Packet Networks"*, John Wiley & Sons LTD, 2001.

[5] William C. Hardy, *"VoIP Service Quality"*, McGraw-Hill, 2003.

[6] Yi J. Liang, N. Fọrber and B. Girod, *"Adaptive playout scheduling using time-scale modification in packet voice communications"* in Proc. IEEE int. Conf. Acoust., Spech, Signal Processing, Vol 3, 2001.

[7] Moo Young Kim and Renat Vafin, *"Packet-loss recovery techniques for VoIP"*, 2001.

[8] Nortel Networks, *"RTP/UDP/IP header removal and construction for optimized voice bearer"*, 2000.

[9] Tomas Andersson, *"Techniques used in Voice over IP Systems"*, 2002.

[10] Nguyễn Quốc Cường, *"Internet working với TCP/IP"*, tập 1, Nhà xuất bản Lao động xã hội, 2001.

[11] Nguyễn Hồng Sơn, *"Kỹ thuật điện thoại qua IP và Internet"*, Nhà xuất bản Lao động xã hội, 2003.

[12] Phùng Văn Vận, *"Điện thoại IP"*, Nhà xuất bản Bưu điện, 2003.

[13] www.ietf.org, (1999), *"RFC 2198, RFC 2733, RFC 1889"*,.

[14] ITU-T recommendations G.7xx

MỤC LỤC

CÔNG NGHỆ VoIP
VÀ CÁC GIẢI PHÁP NÂNG CAO CHẤT LƯỢNG DỊCH VỤ

Chịu trách nhiệm xuất bản
LƯU ĐỨC VĂN

Biên tập : ĐÀO THỊ MINH – TRẦN VŨ THƯỞNG
NGUYỄN VĂN VĨNH
Chế bản : VŨ HỒNG NHUNG
Sửa bản in : NGUYỄN VĂN VĨNH
Trình bày bìa : PHAN THẾ VINH

NHÀ XUẤT BẢN BƯU ĐIỆN

Trụ sở : 18 Nguyễn Du, TP. Hà Nội
Điện thoại : 04-9431283, 04-9432438 Fax: 04-9431285
E-mail : bientap@hn.vnn.vn
Chi nhánh : 27 Nguyễn Bỉnh Khiêm, Quận I, TP. Hồ Chí Minh
Điện thoại : 08-9100925 Fax: 08-9100924
E-mail : chinhanh-nxbbd@hcm.vnn.vn